AFRL-SN-RS-TR-2004-123
Final Technical Report
May 2004

RADAR SIGNAL DETECTION BASED ON BAYESIAN HIERARCHICAL MODELS AND IMAGE ANALYSIS TECHNIQUES

Syracuse University

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-SN-RS-TR-2004-123 has been reviewed and is approved for publication

APPROVED: /s/

PETER A. ZULCH
Project Engineer

FOR THE DIRECTOR: /s/

RICHARD G. SHAUGHNESSY, Lt. Col., USAF
Chief, Rome Operations Office
Sensors Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	MAY 2004	Final May 01 – Aug 03

4. TITLE AND SUBTITLE
RADAR SIGNAL DETECTION BASED ON BAYESIAN HIERARCHICAL MODELS AND IMAGE ANALYSIS TECHNIQUES

5. FUNDING NUMBERS
C - F30602-01-2-0525
PE - 62204F
PR - 762R
TA - SN
WU - P3

6. AUTHOR(S)
Pramod K. Varsheny and Biao Chen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Syracuse University
Department of EE and CE
121 Link Hall
Syracuse New York 13244

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/SNRT
26 Electronic Parkway
Rome New York 13441-4514

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
AFRL-SN-RS-TR-2004-123

11. SUPPLEMENTARY NOTES
AFRL Project Engineer: Peter A. Zulch/SNRT/(315) 330-7861/ Peter.Zulch@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***
The research undertaken under this effort has two major components. First, the application of Bayesian inference theory is applied to problems ranging from Distributed Detection with multiple sensors, clutter scene characterization/identification for airborne radar systems, to adaptive CFAR detection with heterogeneous clutter. Secondly, multichannel radar detection algorithms are developed that are particularly suitable for airborne radar surveillance systems operating in a complex clutter/interference/noise environments.

14. SUBJECT TERMS
Radar Detection, Multi-Sensor Detection, Baysian Inference, Multi-Channel Detection

15. NUMBER OF PAGES
74

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Contents

List of Figures

List of Tables

Chapter 1

Executive Summary

The research project undertaken under this effort has two major components. In the first one, we investigate the Bayesian inference theory and its application to research problems ranging from distributed detection with multiple sensors, clutter scene characterization and clutter patch identification for airborne radar systems, and adaptive CFAR detection with heterogeneous clutters. In the second part, multichannel radar detection algorithms were developed that were particularly suitable for airborne radar surveillence systems operating in a complex clutter/interference/noise environment. Several major tasks have been carried out under these two components, all of them tackle problems that are of great relevance and importance to the United States Air Force. Significant progress has been made in both fronts and the research work has been well recognized as evidenced by publications of the results on peer-reviewed leading journals in the relevant areas. In all, three journal papers have been published with one more paper currently under revision. All of these papers are in the prestigious IEEE Transactions, the most authoritative journals in the respective technical areas. In addition, numerous papers were published and presented in many leading technical conferences. Some of them have already been cited by other researchers working in similar areas.

Chapter 2

Introduction

In this report, we present results obtained from research effort supported by AFRL through Corporative Agreement: # F30602-01-2-0525. There are two major thrusts under this effort: the first one (Chapters 3 through 5) is the application of modern Bayesian inference tools to radar signal processing ranging from multi-sensor detection, to clutter scene identification, to CFAR detection in the presence of clutter-edge heterogeneity. In the second part (Chapters 6 and 7), multi-channel radar detection problem has been addressed using classical detection theory and a new generalized likelihood ratio test statistic that can deal with the presence of both clutter and noise is proposed.

Data fusion and distributed detection have been studied extensively and numerous results have been obtained during the past two decades. In Chapter 3 the design of fusion rule for distributed detection problems is reexamined and a novel approach using Bayesian inference tools is proposed. Specifically, the decision fusion problem is reformulated using hierarchical models and a Gibbs sampler is proposed to perform posterior probability based fusion. Performance wise, it is essentially identical to the optimal likelihood based fusion rule whenever it exists. The true merit of this approach is its applicability to various complex situations, e.g., in dealing with unknown signal/noise statistics where likelihood based fusion rule may not be easy to obtain or may not even exist.

Radar CFAR detection is addressed in Chapter 4. Motivated by the frequently encountered problem of clutter-edge heterogeneity, we model the secondary data as a probability mixture and impose a hierarchical model for the inference problem. A two-stage CFAR detector stucture is proposed. Empirical Bayesian inference is adopted in the first stage for training data selection followed by a CFAR processor using the identified homogeneous training set for target detection. One of the advantages of the proposed algorithm is its inherent adaptivity; i.e., the threshold setting is much less sensitive to the nonstationary environment compared with other standard CFAR procedures.

In Chapter 5, we address the problem of clutter patch identification based on Markov random field (MRF) models. MRF has long been recognized by the image processing community to be an accurate model to describe a variety of image characteristics such as texture. Here, we use the MRF to model clutter patch characteristics, captured by a radar receiver or radar imagery equipment, due to the fact that clutter patches usually occur in connected regions. Furthermore,

we assume that observations inside each clutter patch are homogenous, i.e., observations follow a single probability distribution. We use the Metropolis-Hasting algorithm and the reversible jump Markov chain algorithm to search for solutions based on the Maximum a Posteriori (MAP) criterion. Several examples are provided to illustrate the performance of our algorithm.

Chapter 6 studies multi-channel radar detection in the presence of both Gaussian and non-Gaussian disturbance. We develop maximum likelihood parameter estimates for spherically invariant random processes (SIRP) in the presence of white Gaussian noise. Both cases with known and unknown white noise variance are treated. As the estimators do not admit closed-form solutions, numerical iterative procedures are developed that are guaranteed to at least converge to the local maximum. The developed estimate allows us to construct a generalized likelihood ratio test (GLRT) for the detection of a signal with constant but unknown amplitude embedded in both Gaussian noise and SIRP disturbances. This new GLRT compares favorably to existing detection schemes that neglect the existence of white Gaussian noise.

Compound-Gaussian processes have found important applications in modeling clutter returns for high-resolution radar. In the last chapter we develop a maximum likelihood estimate for the covariance structure of a compound Gaussian process. The performance of the covariance matrix estimator is then evaluated in the context of adaptive radar detection. Through extensive numerical simulation and by using a popular CFAR detector for coherent pulse train detection in non-Gaussian clutter, we show that the proposed estimator provides better detection performance over existing covariance matrix estimators

Chapter 3

A Bayesian Sampling Approach to Decision Fusion

Data fusion and distributed detection have been studied extensively and numerous results have been obtained during the past two decades. In this chapter the design of fusion rule for distributed detection problems is reexamined and a novel approach using Bayesian inference tools is proposed. Specifically, the decision fusion problem is reformulated using hierarchical models and a Gibbs sampler is proposed to perform posterior probability based fusion. Performance wise, it is essentially identical to the optimal likelihood based fusion rule whenever it exists. The true merit of this approach is its applicability to various complex situations, e.g., in dealing with unknown signal/noise statistics where likelihood based fusion rule may not be easy to obtain or may not even exist.

3.1 Introduction

Data fusion refers to the inference problem where data are gathered from distributed agents and are processed collectively at a fusion center. Fig. 3.1 is a simple illustration of a data fusion system, where data generated by some underlying phenomenon and collected at local agents are transmitted, with possible preprocessing, to a central processor where inference about the underlying phenomenon and any ensuing decisions are to be made. Extensive research has been conducted in the past two decades and is documented in [1–3].

The problem of interest in this chapter is distributed detection, and in particular, the fusion of decisions from local sensors. Spurred by many real world problems, many of them related to military surveillance applications, distributed detection has been vigorously studied and many fundamental results have been obtained in this area [1, 2]. Performance wise, it is desirable for the local detectors to send the raw data to the fusion center, where the problem of interest is often termed as pre-detection fusion [4]. Such an approach usually yields optimal detection performance as there is no information loss at the local sensor. However, in practical situations, the sensors are scattered and are often located far away from the fusion center. The information that can be transferred from the sensors to the fusion center is, therefore, limited by the communication

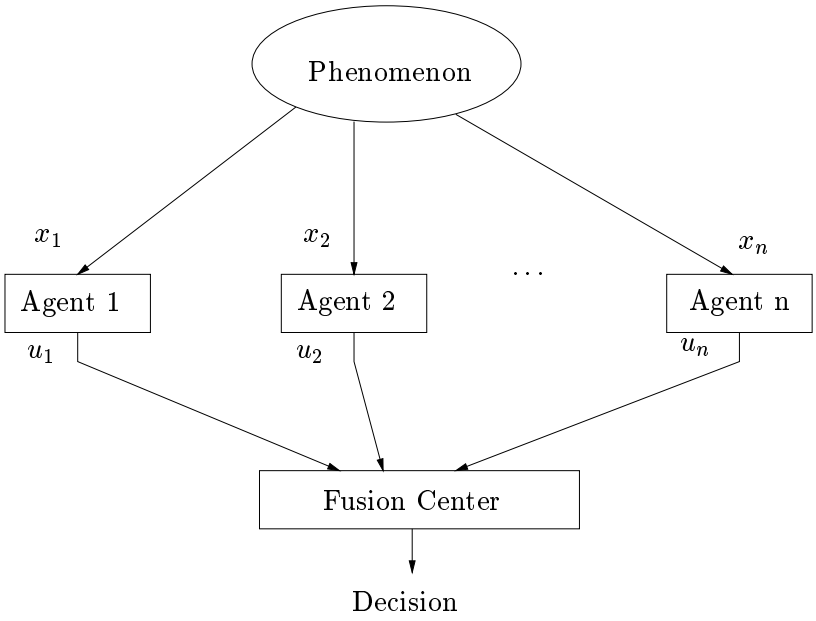


Figure 3.1: Data fusion problem

channel and other practical considerations. These limitations often mandate that the observations at local sensors be processed (compressed) prior to being sent out to the fusion center. The need for local processing greatly complicates the problem. In fact, the majority of the literature deals with the optimization of local decision rules, which is often times coupled with the fusion rule, by applying various classical inference tools, such as the Neyman-Pearson criterion and Bayesian detection theory. The success of classical statistical inference methods, however, depends largely on some simplifying assumptions that are often not valid in practice.

In certain situations, however, it may be desirable to consider the design of fusion rules independent of the local decision rules. For example, it may not always be practical to constantly adapt the local decision rule according to a changing environment. For a set of fixed local decision rules, Chair and Varshney [5] proved that the optimal fusion rule based on the data received from the sensors is a weighted sum of local decisions provided the performance indices (in terms of false alarm rate and probability of detection) of local detectors are available. Similar results exist for soft output (multibit quantization) from local sensors [2]. Without the conditional independence assumption, however, the fusion rule is much more complicated and results are usually limited to the case of hard decisions at the local output [6–8]. More general results about distributed detection for spatially correlated observations are also available, see, e.g., [9]. Notice that the above work requires explicit knowledge of the performance of each local detector which is not possible when the signal and/or noise statistics are not completely known.

We propose in this chapter a novel approach using Bayesian sampling to attack the decision fusion problem. The approach is readily applicable to much more complex situations where the classical approach will either fail or become too complicated to carry out, such as problems involving unknown signal/noise statistics, and possibly sensors with dependent observations. A key

observation in applying Bayesian inference to distributed detection is to recognize the enormous resemblance between hierarchical models and distributed detection problems. Hierarchical models are applicable to problems where parameters and/or observations interact (in the form of conditional probability) through a certain hierarchical structure. Fig. 3.1 clearly suggests that such a hierarchy exists for the data fusion problem. A hierarchical model is "Bayesian friendly" because its structured dependence among variates allows easy calculation of conditional probabilities. Under the Bayesian inference framework, unknown parameters associated with the model are assumed random with a suitable prior (often chosen to be vague, or non informative) and the aim is to obtain the posterior distribution of those parameters that are of interest. Notice here the term 'Bayesian' has a different meaning than that in Bayesian detection theory. In Bayesian detection theory, each different hypothesis is assigned a prior probability which is assumed *fixed and known* and the ensuing inference procedure is classical in essence. In the true Bayesian inference paradigm, all the parameters involved, including the prior probabilities on different hypotheses, are assumed random and the goal is to obtain the posterior probability of every unknown parameter.

3.2 Hierarchical modeling and Gibbs sampler for decision fusion

3.2.1 Hierarchical modeling for binary detection

Hierarchical models (HM) are best suited to describe situations where observations and/or parameters can be related or connected to each other through a hierarchical structure. A well-known application is in problems involving exchangeable parameters/variables — they come from a common population distribution yet their order, or index, does not carry any information [10]. In the sensor fusion problem, for example, the order, or labeling, of the sensors is irrelevant to the inference task at hand. This exchangeability allows us to link the set of parameters together by assuming a common population distribution for them, thus creating the top of the pyramid upon which a hierarchical model can be built. We mention here that although modeling based on exchangeability is a prominent example for the application of hierarchical models, it is not a requisite for HM to be useful.

Specification of a hierarchical model usually involves two steps. The first is to understand the physical phenomenon and the associated dependence structure. This enables us to construct a hierarchical model that reasonably approximates the physical meaning of the underlying problem. The proposed model should incorporate those unknown parameters that are relevant to the inference task and clearly describe the dependence structure among these parameters. This step often involves some necessary approximations — an accurate characterization of a complex system may be either impossible to obtain or too complicated to work with. The second involves selection of the priors for all the random parameters involved in the hierarchical model. While specification of the priors is often considered subjective in the absence of *a priori* information, there are established rules and theory that can serve as a guideline to select a prior to avoid subjectivism [11]. A prominent example of is Jeffreys' rule [12] that is based on the invariance of Fisher information and often

results in priors that are conceptually vague, or *non informative*.

In reference to Fig. 3.1, we see that the local decisions u_i's are functions of local observations x_i's, which in turn come from distributions dictated by a *common* underlying phenomenon. This is precisely the type of hierarchical structure that we discussed earlier where parameters/observations interact with each other in the form of conditional probabilities through a layered structure. Therefore we notice a natural relationship between decision fusion and hierarchical models. This is further elaborated below.

We define, for binary hypothesis testing, a random variable Z:

$$Z = \begin{cases} 0 & \text{if } H_0 \text{ is true.} \\ 1 & \text{if } H_1 \text{ is true.} \end{cases}$$

The inference on the underlying hypothesis is now converted to inference on the random variable Z. Under the Bayesian inference framework, Z is most commonly assumed to be a Bernoulli random variable with success probability θ. Notice the subtle difference from Bayesian detection theory: In Bayesian detection theory, Z is assigned a fixed prior on each hypothesis, which may be hard to determine or justify in practice. Here, the success probability θ is not fixed, but rather is assigned a prior that makes it more robust in practice. For conjugacy*, we assume a *beta* prior for θ. We make a note here that the uniform distribution, which is conceptually the most vague prior is a special case of *beta* distribution.

Given Z, the underlying hypothesis, the local observations X_i's can be specified by conditional probabilities depending on the value of Z. The last layer of the hierarchy consists of the local decision rules, which are assumed to be fixed here, and whose outputs, denoted by $\mathbf{U} = (U_1, \cdots, U_n)$ (c.f. Fig. 3.2), are the observations at the fusion center where a final decision is to be made. The joint distribution of all the parameters involved, under the hierarchical model assumption, is easily obtained as

$$f(\theta, Z, \mathbf{X}, \mathbf{U}) = f(\theta) P(Z|\theta) f(\mathbf{X}|Z) I_{\mathbf{U}=g(\mathbf{X})} \tag{3.1}$$

where $I_{\mathbf{U}=g(\mathbf{X})}$ is the indicator function defined as

$$I_{\mathbf{U}=g(\mathbf{X})} = \begin{cases} 1 & \text{If } \mathbf{U} = g(\mathbf{X}). \\ 0 & \text{Otherwise.} \end{cases}$$

The hierarchical model for the decision fusion problem is shown in Fig. 3.2. We describe each layer in the hierarchical model as follows.

- Prior on θ — For conjugacy, it is most convenient to choose the prior for θ as $beta(\alpha, \beta)$ distribution. The resulting posterior of θ given Z is then also a *beta* distribution given

*Conjugacy refers to the situation where prior and posterior follow the same type of distribution. For example, in the current situation, given that the likelihood (probability distribution of Z given θ) is a Bernoulli random variable, *beta* prior on θ will result in a posterior probability of θ also in the form of *beta* distribution with different parameters. Although conjugacy is not necessary in theory for the Bayesian inference framework to work, it nonetheless is more convenient to work with and often times proves to be very robust. Adding a hyper prior layer may further mitigate the concerns about robustness.

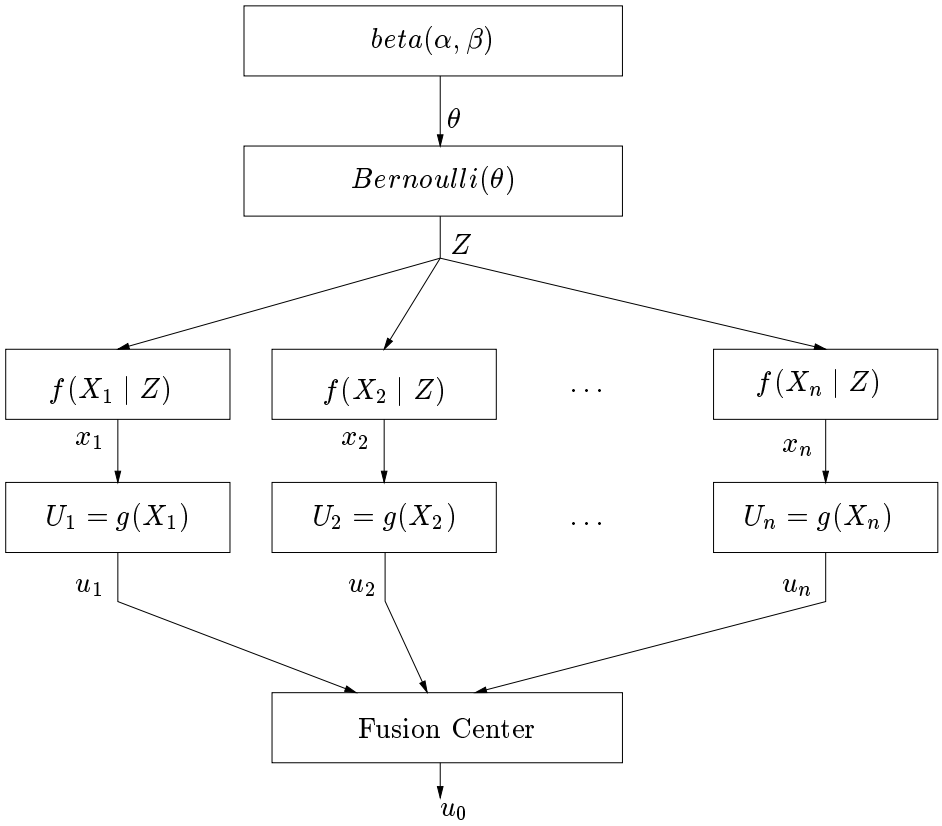


Figure 3.2: Reformulation of the distributed detection problem using a hierarchical model

that Z is a Bernoulli random variable. For simplicity and non-informativeness, we choose $\alpha = \beta = 1$ in our simulations which reduces the *beta* distribution to a uniform distribution. An alternative approach is to choose hyper priors for α, β to allow more modeling flexibility and increased robustness [10, 13].

- First level — Z is assumed a Bernoulli random variable with success probability θ.
- Second level — It represents the likelihood function. Given Z, we should be able to calibrate probabilistically the input to the local sensors in terms of $f(\mathbf{X}|Z)$ where $\mathbf{X}$ includes observations at all sensors. Notice here that there is no restriction as to the form of the likelihood. Therefore, it should be possible to extend this framework to problems involving correlated observations among local sensors.
- Third level — It represents the local decision stage. A deterministic mapping (quantization) from the observation to local decision $U_i = g(X_i)$ is assumed known. The local decision vector $\mathbf{U} = [U_1, \cdots, U_n]$ forms the input to the fusion center.

In the present work, we assume conditional independence among different sensor observations. This assumption simplifies our presentation and is implicitly used in Figure 3.2. Our goal is to infer about the Bernoulli random variable Z using its posterior probability. From the model, the

posterior probability based inference on Z is equivalent to estimating the posterior mean for θ, which is precisely the definition of probability of Z being 1.

3.2.2 A Gibbs sampler for decision fusion

Bayesian inferencing aims at finding the posterior probability of the parameters of interest. Analytical results are most desirable if they exist. In complex situations where a large number of parameters are involved, as is the case with hierarchical models, obtaining explicit analytical results may not be practical. A powerful tool for these problems is the Bayesian sampling approach that we employ here.

Bayesian sampling approach carries out the inference process by finding the empirical posterior distribution of the parameters of interest. Under this paradigm, samples of these random parameters are generated that follow the posterior distribution of the parameters conditioned on the observations. Sampling based approach has long been recognized as a key to the success of Bayesian inference methods due to the fact that analytical solutions (even analytical approximations) often fail in complex situations. Various posterior simulation methods have been developed for Bayesian inferencing, including direct simulation and successive approximation, among others, see [13]. Although they have attained some success in applications, the direct sampling approach has encountered enormous difficulty when dealing with complex problems, e.g., problems with high dimensional parameter sets. This problem is largely solved with the discovery (or more precisely, the rediscovery) of Markov Chain Monte Carlo (MCMC) methods which are mainly responsible for the revival and popularity of the Bayesian inference approach since the late eighties [14]. MCMC circumvents the difficulty of the direct approach by carefully designing a sample trajectory that, upon convergence, assumes a stationary distribution that follows the desired posterior distribution.

For hierarchical models, Gibbs sampler has proved to be the most efficient and effective among various MCMC methods. Gibbs sampler, in short, is an iterative sampling scheme where at each iteration, parameters are sampled alternatively using their full conditional distributions. To illustrate this, assume that we have obtained a joint posterior distribution of all the parameters involved: $f(\gamma_1, \cdots, \gamma_n|\mathbf{y})$ where $\mathbf{y}$ represents the observations. From the joint posterior, we can obtain the following full conditional distributions for each parameter:

$$\begin{gathered} f(\gamma_1 \mid \gamma_2, \gamma_3, \cdots, \gamma_n, \mathbf{y}) \\ f(\gamma_2 \mid \gamma_1, \gamma_3, \cdots, \gamma_n, \mathbf{y}) \\ \vdots \\ f(\gamma_n \mid \gamma_1, \gamma_2, \cdots, \gamma_{n-1}, \mathbf{y}) \end{gathered}$$

If we draw samples iteratively according to the above full conditional distributions, then it can be shown that, under certain regularity conditions, the sample sequence will converge to a distribution specified by the joint posterior probability of all the parameters [15]. Notice that, in general, each γ_i in the above full conditional distribution may contain a subset of the parameters. For a hierarchical model, conditioned on any particular layer, the variables that belong to the layers above and below

are independent of each other. Therefore the full conditional distributions often reduce to very simple and low (and often times one) dimensional form that is easy to sample. In the following, a Gibbs sampler is presented for the hierarchical model for decision fusion described in the previous section.

- θ

$$f(\theta|Z,\mathbf{X},\mathbf{U}) = f(\theta|Z) = \frac{f(\theta,Z)}{P(Z)} = \frac{f(\theta)P(Z|\theta)}{\int f(\theta)P(Z|\theta)d\theta} \tag{3.2}$$

  where under the Bernoulli assumption on Z and *beta* prior for θ, the above probability is also a *beta* random variate — the so-called conjugacy property.

- $\mathbf{X}$

$$f(\mathbf{X}|\theta,Z,\mathbf{U}) = f(\mathbf{X}|Z,\mathbf{U}) = \frac{f(\mathbf{X}|Z)I_{g(\mathbf{X})=U}}{\int_{g(\mathbf{X})=\mathbf{U}} f(\mathbf{x}|z)}d\mathbf{x}$$

  This follows from the fact that $\mathbf{U}$ is a deterministic function of the local observation vector $\mathbf{X}$. In the case that $g(\cdot)$ is a threshold rule, the above conditional probability results in a truncated likelihood function at local sensors.

- Z

$$f(Z|\theta,\mathbf{X},\mathbf{U}) = f(Z|\theta,\mathbf{X}) = \frac{f(\theta,Z,\mathbf{X})}{f(\theta,\mathbf{X})} = \frac{f(\theta,Z,\mathbf{X})}{\sum_Z f(\theta,Z,\mathbf{X})} \tag{3.3}$$

  where $f(\theta,Z,\mathbf{X}) = f(\theta)P(Z|\theta)f(\mathbf{X}|Z)$. The above follows from the fact that given $\mathbf{X}$, $\mathbf{U}$ and Z are independent of each other, a direct result of the hierarchical structure.

The Bayesian sampling based fusion rule can be summarized as follows.

Initialization The unknown variables involved are θ, $\mathbf{X}$ and Z and the observations are contained in $\mathbf{U}$. Initial values for the unknown parameters may be chosen arbitrarily. Multiple sampling paths can be obtained by choosing different initial values for the unknowns. This would facilitate the monitoring of the convergence of the Gibbs samplers as mentioned later. Notice that the initial values for $\mathbf{X}$ must be chosen such that $\mathbf{U} = g(\mathbf{X})$, i.e., they need to be consistent with the sensor output vector according to the local decision rule. We denote the initial values by θ^0, $\mathbf{X}^0$, and Z^0.

Bayesian sampling This is the step to generate samples for the unknowns that, after convergence, should follow their respective posterior probability given observations $\mathbf{U}$.

- Given θ^t, $\mathbf{X}^t$, and Z^t, we generate θ^{t+1}, $\mathbf{X}^{t+1}$, and Z^{t+1} using the random variable generator specified by the full conditional distributions in (3.2) through (3.3). Notice the specification of these full conditional distributions require both the observation $\mathbf{U}$ and the samples at the previous instant, i.e., θ^t, $\mathbf{X}^t$, and Z^t.
- Increment t and repeat the above sampling procedure until we have obtained T samples where T is a preselected number that is large enough to guarantee the convergence of the sample path.

Inference This is the step to obtain the final fusion result. Notice that the previous sampling step has generated a sequence of samples for θ, that is, $\theta^0, \theta^1, \cdots, \theta^T$, which are assumed to follow the posterior distribution for θ. Therefore we use the following decision rule

$$Z = \begin{cases} 1 & \bar{\theta} \geq 0.5 \\ 0 & \bar{\theta} < 0.5 \end{cases}$$

where $\bar{\theta}$ is the mean value of the samples for θ, i.e.,

$$\bar{\theta} = \frac{1}{T - t_0} \sum_{t=t_0+1}^{T} \theta^t$$

where the first t_0 samples are excluded as they are considered to be in the transient phase of the Gibbs sampler.

Notice that the parameter θ is defined as the probability that $Z = 1$, the above decision fusion rule is essentially the posterior probability fusion rule using the Bayesian sampling approach.

Some discussion regarding the convergence of the Gibbs sampler are in order. While Markov process theory [15] guarantees the convergence of a Gibbs sampler, it does not tell exactly when the convergence will occur. Nonetheless, there are many heuristic ways to check the convergence of a Gibbs sampler [13, Chapter 11]. For example, one can run multiple Gibbs samplers with different starting points and monitor the individual sample trajectory. Convergence occurs if all trajectories start to 'merge' with each other. Quantitative convergence measure can also be adopted for convergence check. For example, if multiple Gibbs samplers are available, the variances of between- and within-sequence samples of some particular scalar parameters can be used. In this chapter, since we care mostly about the feasibility of the methodology and performance of the new approach, we would simply preselect a fixed and large enough iteration number T which can be determined empirically by simple numerical experiments. Depending on the dimension of the problem, we may choose smaller T to reduce the computational complexity in real time applications.

A nice property of the Bayes sampling approach is its *plug-in* capability — as long as the probability distribution of local observations and the local decision rules are well defined, we can simply plug them in the above formulation and crank up the Gibbs sampler. Further, Bayesian inference approach is advantageous in dealing with signal/noise uncertainties. In the presence of unknown signal/noise statistics, likelihood based fusion rules have to resort to various means, often times *ad hoc*, to deal with the so-called nuisance parameters. This problem, however, is essentially nonexistent in the Bayesian inference framework — every unknown parameter is assumed random with a suitable prior, including the nuisance parameters.

3.2.3 Extension to multiple hypothesis testing

We define, for M-ary hypothesis testing, a random variable Z:

$$Z = \begin{cases} 1 & \text{if } H_1 \text{ is true.} \\ \vdots & \vdots \\ M & \text{if } H_M \text{ is true.} \end{cases}$$

The inference on the underlying hypothesis is now converted to inference on the random variate Z. We assign each hypothesis a prior probability p_i (with $\sum_i p_i = 1$) — Z is therefore a single trial multinomial random variable with success probabilities $(p_1, \cdots, p_{M-1}) \stackrel{\triangle}{=} \theta^{\dagger}$. We note here that a multinomial random variable is a high dimensional generalization of the binomial random variable of which Bernoulli is a special case with number of trials equal to one. Under *Bayesian detection theory*, each p_i is assigned a fixed value. Under the Bayesian inference framework, those p_i's are now assumed random. A clear advantage of the randomness assumption is its robustness against possible prior mismatch. For the set of p_i's, we further assign a Dirichlet prior for conjugacy [16]. Again, this is analogous to the binary detection case where a *beta* prior is used for the Bernoulli variable Z — Dirichlet random variable is a high dimensional generalization for the *beta* random variable which is a natural choice for describing the distribution for *probabilities*. Given this definition, the joint distribution of all the parameters involved has the formal representation as in (3.1).

The above hierarchical model is similar to that described in Section 3.2.1 and is detailed below.

- Prior on $\theta = (p_1, \cdots, p_{M-1})$ — Dirichlet with parameters $\alpha_1, \cdots, \alpha_M$. Thus

$$f(\theta) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_M)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_M)} p_1^{\alpha_1 - 1} \cdots p_{M-1}^{\alpha_{M-1} - 1} \left(1 - \sum_{i=1}^{M-1} p_i\right)^{\alpha_M - 1}$$

  where $\Gamma(\cdot)$ is the gamma function defined, for real z, as [17, page 942]

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt$$

  For simplicity and non informativeness, we can choose $\alpha_1 = \cdots = \alpha_M = \alpha$ where α is close to zero [13]. As before, an alternative approach is to choose hyper priors for α_i's to allow more modeling flexibility but it may complicate the computation to a certain extent. Again, we emphasize that Dirichlet prior is a high dimensional generalization of the *beta* distribution which is usually used to model the prior for *probabilities*.

- First level — $Z \sim Multinomial(1, \theta)$. This is a high dimensional generalization of a Bernoulli trial.

- Second level — $f(\mathbf{X}|Z)$, the likelihood function. Notice here that the likelihood can be different from sensor to sensor.

- Third level — $U_i = g(X_i)$, local decision rule.

This hierarchical model for the multiple hypothesis testing case can be used for the fusion of classifiers.

[†]Because of the constraint $\sum_i p_i = 1$, there are only $M-1$ free parameters in the density function for multinomial random variables.

3.3 Examples

In this section we present a number of examples to demonstrate the performance of the Bayesian sampling based fusion rule. Our purpose is two-fold: First we want to show that its performance is virtually identical to that of the optimal likelihood based fusion rule whenever it exists. Second, we want to demonstrate that the new fusion method works well in situations where the likelihood based fusion rule does not exist. For the latter, we choose the scenario where noise uncertainty exists in terms of unknown noise variance. We present examples both for binary hypothesis testing and multiple hypothesis testing.

An important issue in the implementation of the Gibbs sampler is the convergence of the sample sequence, i.e., at what point can we claim that the samples start to follow the posterior probability. Various means can be used to determine the approximate convergence including monitoring the convergence of certain scalar parameters. For example, we can use several independent Gibbs samplers with possibly different prior specifications. Approximate convergence occurs when different sample trajectories start to become 'indistinguishable'. In our simulation, we use 250 iterations in the Gibbs sampler and choose the last 200 to calculate the posterior mean of the parameters of interest, thus preventing the effect of initial transient phase of the Gibbs sampler. The implication is that convergence occurs usually after 50 iterations. A total of $50,000$ Monte Carlo runs are used in each of the following examples.

3.3.1 Binary detection

Three examples are presented. The first two consider the testing of possible shift in the mean under different noise statistics, i.e., the problem of detecting a constant amplitude signal observed in noise. The last example deals with uncertainty in noise variance under the Gaussian assumption.

Example 1 — Gaussian shift in mean

The first example we study is a simple shift in the mean problem under Gaussian noise. Under H_0, each local observation is assumed to be zero mean Gaussian with variance σ^2 while under H_1 the local observation has a nonzero mean μ and is otherwise the same as under H_0. Further we assume that both μ and σ^2 are known. As for local decisions, we assume that a simple thresholding with threshold $\tau = \mu/2$ is used, i.e.,

$$U_i = g(X_i) = \begin{cases} 0 & X_i < \tau \\ 1 & X_i \geq \tau \end{cases} \tag{3.4}$$

Since the observations at the local sensors are independent identically distributed, it is easily seen that the optimal likelihood based fusion rule relies solely on the sum of u_i's [5].

To develop a Gibbs sampler for this problem, we need to first find the joint posterior distribution of all the parameters. This can be readily written as

$$\begin{aligned} f(\theta, Z, \mathbf{X} \mid \mathbf{U}) &\propto f(\theta, Z, \mathbf{X}, \mathbf{U}) \\ &= f(\boldsymbol{\theta}) P(Z|\theta) f(\mathbf{X}|Z) I_{\mathbf{U}=g(\mathbf{X})} \end{aligned}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{z}(1-\theta)^{1-z} \prod_{i=1}^{n}\left[z\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu)^2/2\sigma^2} + (1-z)\frac{1}{\sqrt{2\pi}\sigma}e^{-x_i^2/2\sigma^2}\right]\prod_{i=1}^{n} I_{U_i=g(X_i)} \quad (3.5)$$

Now we need the full conditional distribution of each random parameter involved. Using the hierarchical model specified in Section 3.2, we have, for each unknown parameter,

- θ

$$f(\theta \mid Z) = \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+z)\Gamma(\beta-z+1)}\theta^{\alpha+z-1}(1-\theta)^{\beta-z}$$

  i.e., it is $beta(\alpha+Z, \beta-Z+1)$. This is so because it is easy to identify that the conditional probability must be proportional to $\theta^{\alpha+Z-1}(1-\theta)^{\beta-Z}$.

- X_i

$$f(X_i \mid Z, U_i) = \frac{\left[z\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu)^2/2\sigma^2} + (1-z)\frac{1}{\sqrt{2\pi}\sigma}e^{-x_i^2/2\sigma^2}\right] I_{u_i=g(x_i)}}{\int_{u_i=g(x_i)}\left[z\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu)^2/2\sigma^2} + (1-z)\frac{1}{\sqrt{2\pi}\sigma}e^{-x_i^2/2\sigma^2}\right]dx_i}$$

- Z

$$P(Z \mid \theta, \mathbf{X}) = \frac{\theta^z(1-\theta)^{1-z}\prod_{i=1}^{n}\left[z\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu)^2/2\sigma^2} + (1-z)\frac{1}{\sqrt{2\pi}\sigma}e^{-x_i^2/2\sigma^2}\right]}{\sum_z \theta^z(1-\theta)^{1-Z}\prod_{i=1}^{n}\left[z\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu)^2/2\sigma^2} + (1-z)\frac{1}{\sqrt{2\pi}\sigma}e^{-x_i^2/2\sigma^2}\right]}$$

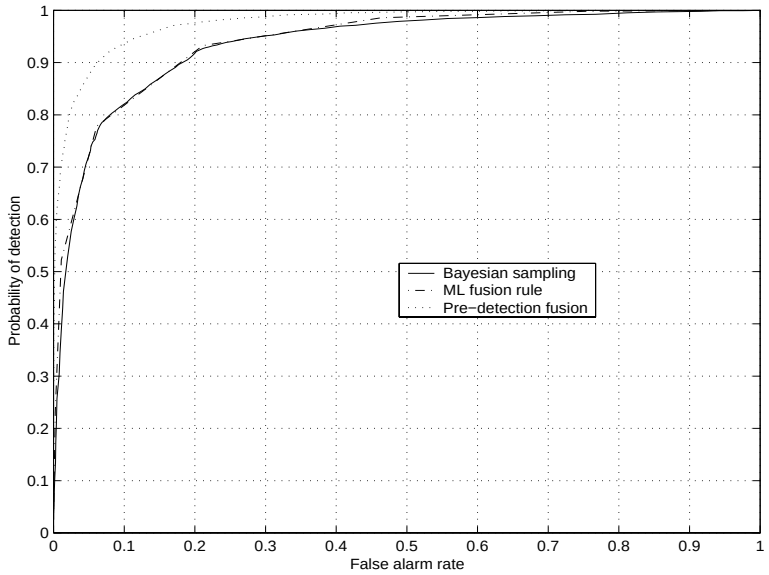


Figure 3.3: ROC curves for different fusion rules for simple Gaussian shift in mean problem with unit variance. Here the number of sensors is 8 and the mean shift is 1. The threshold used at each local detector is 0.5.

In Fig. 3.3, we provide the simulation results on the performance of different fusion rules. Eight sensors are assumed. The shift in the mean is assumed equal to one with unity noise variance. The threshold at the local detectors is 0.5. We see that the Bayesian sampling approach has virtually the same performance as the optimal likelihood based fusion rule, which in this case is simply the thresholding of the sum of sensor outputs U_i because of the symmetry among the sensors [5]. Also plotted for reference is the ideal ROC curve of pre-detection fusion — the raw observations at local sensors are assumed available at the fusion center where likelihood based detection is performed.

While performance wise the Bayesian sampling approach does not have any advantage over the likelihood based fusion rule, and in this particular case, its implementation is much more involved than the likelihood based fusion rule, we should emphasize here that the true merit of the new approach is its wide applicability to complex situations where the classical approach based on conventional detection theory suffers performance degradation or fails to apply.

Example 2 — Different sensor noise statistics

In this example, we investigate the case where different sensors experience different noises. We choose Gaussian and Laplace (double exponential) as the two noise distributions, and in particular, we assume that half of the sensors experience Laplace noise while the other half observe Gaussian noise. Under H_0, the observations are assumed to have zero mean at the local sensors while under H_1, the observations have mean μ. The scale parameter for Laplacian noise and variance for Gaussian noise are all assumed known. Again, local decision rules are assumed to be a simple threshold device.

The likelihood based fusion rule employs a weighted sum of local decisions where the weights depend on the performance indices (false alarm rate and probability of detection) at the local detectors. In the case of heterogeneous noise statistics, the weights corresponding to different sensors will be different even if the thresholds at local sensors are chosen to be the same. Consequently, different threshold values will result in a different set of weights. The Bayesian sampling approach, on the other hand, is essentially identical to the previous example except that the Gaussian density function is replaced with Laplace density for half of the observations. Fig. 3.4 shows the simulation results where $\mu = 1$ and the scale parameter for Laplacian noise and variance for Gaussian noise are all chosen to be 1. The number of sensors is 8, hence 4 of them observe Gaussian noise while the other 4 observe Laplacian noise. The threshold is chosen again to be $\mu/2$ and for this threshold, the likelihood based decision statistic, derived in Appendix A, turns out to be

$$T(\mathbf{u}) = n_1 a + n_2 b + n_3 c + n_4 d \tag{3.6}$$

where

- n_1 is the number of sensors that experience Gaussian noise and declare 1;
- n_2 is the number of sensors that experience Gaussian noise and declare 0;
- n_3 is the number of sensors that experience Laplacian noise and declare 1;
- n_4 is the number of sensors that experience Laplacian noise and declare 0;

- $a = \frac{1-\alpha}{\alpha}$ where $\alpha = P[\mathcal{N}(0,1) > 0.5] \approx 0.3085$;
- $b = \frac{1}{a}$;
- $c = \frac{2-e^{-0.5}}{e^{-0.5}} \approx 1.5681$;
- $d = \frac{1}{c}$.

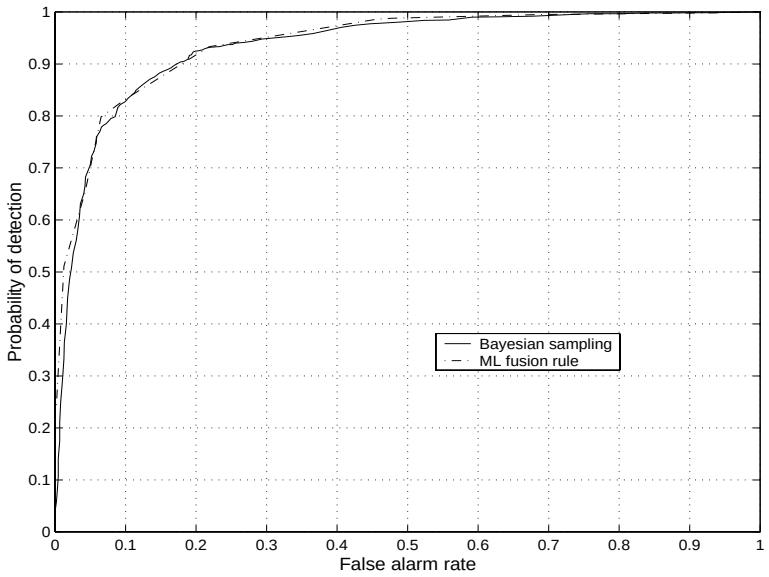


Figure 3.4: ROC curves for likelihood based fusion and Bayesian sampling fusion rules for heterogeneous noise statistics. Here the number of sensors is 8 and the mean shift is 1. Four out of the 8 sensors observe Gaussian noise while the other 4 sensors observe Laplacian noise. The threshold used at each local detector is 0.5.

Clearly from Fig. 3.4, the performance of the Bayesian sampling approach is again virtually indistinguishable from that of the optimal likelihood based fusion rule.

Example 3 — Multibit local decisions with unknown noise statistics

In the previous examples the local detector makes a simple binary decision which results in a significant simplification in terms of the likelihood based fusion rule. In situations where multibit ('soft') decisions are available at the local detectors, the optimal likelihood based fusion rule is more involved. Further, in the previous examples all the statistics of the observations at the local sensors were assumed known — this is why likelihood based based fusion rule can be obtained in a straightforward manner. In practice, however, signal and/or noise statistics may not be available (e.g., they may be time varying). Under this scenario, the likelihood based based scheme does not apply directly. In fact, because of the limited information available at the fusion center (decentralized and truncated sources), even a generalized likelihood ratio based scheme may not be easy to obtain. We also note here that if binary decisions are made at the local detectors, then the uniformly most powerful test exists for this problem even if the noise variance is unknown due to the monotonicity of the likelihood function at the local sensors.

Under the Bayesian inferencing framework, however, there is no substantial difficulty in fusing soft decisions in the presence of signal/noise uncertainty. What needs to be done is to find suitable priors for the unknown *nuisance* parameters and incorporate them in the hierarchical model as well as the Gibbs sampler. Here we use the simple example considered in Example 1 to illustrate the approach. Again, consider the Gaussian shift in mean except that we assume here the local decision yields 2 bits per observation and the noise variance σ^2 is unknown. Specifically, assume the local decision is a simple quaternary quantizer with thresholds $-0.5, 0.5, 1.5$, i.e.,

$$U_i = \begin{cases} 0 & X_i < -0.5 \\ 1 & -0.5 \le X_i < 0.5 \\ 2 & 0.5 \le X_i < 1.5 \\ 3 & X_i \ge 1.5 \end{cases}$$

For the unknown variance σ^2, we choose non informative prior [13], i.e.,

$$f(\sigma^2) \propto \frac{1}{\sigma^2} v(\sigma^2) \tag{3.7}$$

where $v(\cdot)$ is the unit step function[‡]. To facilitate the Gibbs sampler, we need to find the full conditional distribution for σ^2 given the specified prior, and we get

$$\begin{aligned} f(\sigma^2|\mathbf{X}, Z) &\propto (\sigma^2)^{-(\frac{n}{2}+1)} \prod_{i=1}^{n} \left[z e^{-(x_i-\mu)^2/2\sigma^2} + (1-z) e^{-x_i^2/2\sigma^2} \right] \\ &= (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{1}{\sigma^2}\left(\frac{1}{2}\sum_{i=1}^{n}(x_i - z\mu)^2\right)} \end{aligned}$$

which is $inverse-gamma(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(x_i - z\mu)^2)$. An alternative approach is to choose a conjugate prior, that is, we choose the prior for σ^2 to be inverse-gamma which results in the same form of posterior for σ^2. For the Gibbs sampler, all we need is to insert the above full conditional distribution into the iterative sampling scheme described before. We compare the performance to the likelihood based fusion rule assuming perfect knowledge of σ^2. The results are plotted in Fig. 3.5. Clearly, the Bayesian sampling approach is fairly close to the optimal fusion rule even when the noise variance is not known.

3.3.2 Mutliple hypotheses

Two examples involving multiple hypothesis testing, and in particular, ternary hypothesis testing, are given. The first is similar to the binary Gaussian shift in mean example while the second involves unknown noise variance.

[‡]Notice that this prior is improper (non-integrable over $(0, \infty)$), therefore it is necessary to check the properness of the resulting posterior [13]. It turns out however, that the properness is not guaranteed for all possible output values of $\mathbf{U}$. A trivial modification would be to simply truncate the prior, say, we use

$$f(\sigma^2) \propto \frac{1}{\sigma^2} v(\sigma^2 - \sigma_0^2)$$

where σ_0^2 is a small yet positive number. For convenience, we will proceed with the prior in (3.7) as it leads to easy full conditional probability in our presentation.

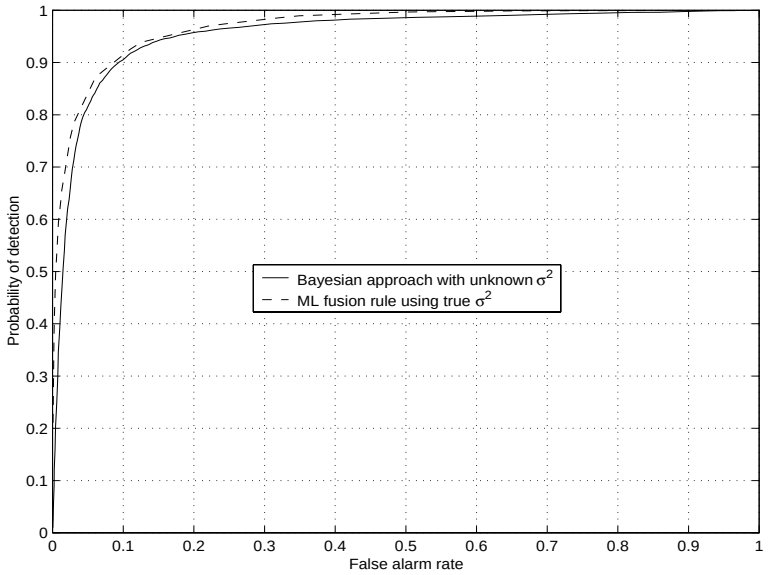


Figure 3.5: ROC curves for various fusion rules for the Gaussian shift in mean problem with unit variance, assumed unknown at the fusion center. Each local detector provides quaternary quantization to the fusion center. The true noise variance is assumed unknown to the Bayesian sampling approach but is assumed known to the likelihood based fusion rule.

Example 4 — Gaussian shift in mean

We start with a simple Gaussian-shift-in-mean example — each hypothesis corresponds to one of three possible mean values of the observations at local sensors which are otherwise assumed Gaussian with known variance. The joint posterior distribution of all the parameters can be readily written as

$$\begin{aligned} f(\theta, Z=m, \mathbf{X} \mid \mathbf{U}) &\propto f(\theta, Z=m, \mathbf{X}, \mathbf{U}) \\ &= f(\theta) P(Z=m|\theta) f(\mathbf{X}|Z=m) I_{\mathbf{U}=g(\mathbf{X})} \\ &= \left[\frac{\Gamma(\alpha_1 + \cdots + \alpha_M)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_M)} p_1^{\alpha_1 - 1} \cdots p_{M-1}^{\alpha_{M-1}-1} \left(1 - \sum_{i=1}^{M-1} p_i\right)^{\alpha_M - 1} \right] p_m \\ &\quad \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i - \mu_m)^2 / 2\sigma^2} \right] \prod_{i=1}^{n} I_{u_i = g(x_i)} \end{aligned}$$

From this we can derive the full conditional distribution of each parameter as follows

- $\theta = (p_1, \cdots, p_{M-1})$

$$f(\theta \mid Z=m) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_M + 1)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_m + 1) \cdots \Gamma(\alpha_M)} p_1^{\alpha_1 - 1} \cdots p_m^{\alpha_m} \cdots p_{M-1}^{\alpha_{M-1}-1} \left(1 - \sum_{i=1}^{M-1} p_i\right)^{\alpha_M - 1}$$

  i.e., it is $Dirichlet(\alpha_1, \cdots, \alpha_m + 1, \cdots, \alpha_M)$.

- X_i

$$f(X_i \mid Z = m, U_i) = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu_m)^2/2\sigma^2} I_{U_i=g(X_i)}}{\int_{U_i=g(X_i)} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu_m)^2/2\sigma^2}\right] dy}$$

- Z

$$P(Z = m \mid \theta, \mathbf{X}) = \frac{p_m \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu_m)^2/2\sigma^2}}{\sum_k p_k \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu_k)^2/2\sigma^2}}$$

For simplicity we choose $\sigma^2 = 1$ and $M = 3$ (ternary hypotheses) with $\mu_1 = -1$, $\mu_2 = 0$, and $\mu_3 = 1$. The total number of sensors is $n = 4$. The local decision rule in this case is also assumed to be a simple ternary quantization rule with quantization thresholds at -0.5 and 0.5. The final decision (classification) is based on the samples of the posterior probability for $(p_1, \cdots, p_M)$ (with $\sum p_m = 1$), and in particular we use the *maximum a posteriori* probability decision rule, i.e., we choose $Z = m$ if the sample mean of p_m is the largest. Performance evaluation is conducted by simulation. Notice for this simple example, maximum likelihood based fusion rule can be easily obtained and we skip the details. The results are summarized in Table 1 where each entry is the classification error probability under each hypothesis. Clearly from the table, the Bayesian sampling scheme is fairly close to the performance of the likelihood based fusion rule.

True Hypothesis	Bayesian	Likelihood Based
H_1 $(\mu = -1)$	0.1911	0.1982
H_2 $(\mu = 0)$	0.4168	0.3529
H_3 $(\mu = 1)$	0.1856	0.1923

Table 3.1: Classification error probability using Bayesian sampling and likelihood based approaches.

Example 5 — Unknown noise variance

In this example, we consider the multiple hypothesis testing problem in the presence of unknown noise variance. Again, consider an almost identical ternary hypothesis testing problem as above except that we assume here the noise variance σ^2 is unknown and the total number of sensors is $n = 8$. We choose a non informative prior for the variance σ^2 as in Section 3.3.1, i.e., $f(\sigma^2) \propto \frac{1}{\sigma^2} v(\sigma^2)$. For this prior, the posterior is found to be

$$f(\sigma^2|\mathbf{X}, Z = m) \propto (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{1}{\sigma^2}\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_m)^2\right)}$$

which is $inverse - gamma(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu_m)^2)$. We compare the performance of the Bayesian sampling based fusion rule with the likelihood based approach with possible parameter (noise variance) mismatch. In this example, the true underlying noise variance is assumed to be unity. In Table 3.2, unity variance is assumed for results presented in rows 1 and 2, while rows 3 and 4 correspond to the likelihood based approach with variance mismatch (σ^2 is assumed to be 4 and 1/4 respectively). From Table 2, the performance of the Bayesian sampling approach (first row)

is a lot closer to that of the true likelihood based fusion with perfect knowledge of noise variance than the likelihood based approaches with parameter mismatch.

True Hypothesis	H_1	H_2	H_3
Bayesian	0.1023	0.3024	0.1041
likelihood based with $\hat{\sigma}^2 = 1$	0.0818	0.2601	0.0820
likelihood based with $\hat{\sigma}^2 = 4$	0.0307	0.5027	0.0300
likelihood based with $\hat{\sigma}^2 = 1/4$	0.1844	0.1135	0.1866

Table 3.2: Classification error using Bayesian sampling and likelihood based approach with possible parameter mismatch. The true noise variance is $\sigma^2 = 1$.

It is also interesting to see how the Bayesian sampling approach performs when there is a fixed prior on the hypotheses as assumed in Bayesian detection theory. Here we assume the prior on the ternary hypotheses to be 0.65, 0.25, and 0.15 respectively. For this case, we know that the optimal fusion rule is the maximum posterior probability decision rule that minimizes the classification error probability. We compare the performance of the Bayesian sampling approach with Bayesian detection theory *using true prior* but with possible noise variance mismatch as it is not known at the receiver. Clearly, from Table 3 where the classification error probabilities for different methods are listed, we see while the performance of the Bayesian sampling approach is inferior to the optimal Bayesian Detection Theory (BDT) using the true noise variance, it certainly is superior to the BDT using mismatched noise variance.

Bayesian sampling	0.1548
BDT with $\hat{\sigma}^2 = 1$	0.1231
BDT with $\hat{\sigma}^2 = 4$	0.1886
BDT with $\hat{\sigma}^2 = 1/4$	0.1676

Table 3.3: Classification error probability using Bayesian sampling and Bayesian detection theory with possible parameter mismatch.

3.4 Discussion

In this chapter, we proposed a Bayesian sampling approach for decision fusion. To facilitate the use of Bayesian inference methodology, a hierarchical model was used to reformulate the distributed detection problem. A Gibbs sampler was designed to obtain the samples of the desired parameters that follow the posterior probability. The ensuing decision fusion is based on the posterior samples generated using the Gibbs sampler. Compared with conventional likelihood based fusion rule, the approach has the following advantages

- Robustness to prior probability assignment. Unlike Bayesian detection theory, we do not need to assign a specific prior on each hypothesis as they may not be available in practice.

- Plug in capability — it is equally applicable to different situations encountered in decision fusion. We expect, from the formulation of the hierarchical model, that the case of correlated sensor observations can be dealt with under the same framework and will be addressed in future work. Indeed, assuming that the correlation structure among sensors are given, we know from 3.1 that $f(\mathbf{X}|Z)$ can be perfectly specified. Therefore, at least conceptually, it is feasible to extend the results to cases involving correlated observations.

- In particular, the approach can be easily adapted to situations where likelihood based fusion does not apply. For example, it is straightforward to deal with distributed detection with unknown signal/noise statistics.

Notice that the situation involving unknown signal/noise statistics can also be dealt with within the classical likelihood based inference framework. For example, generalized likelihood ratio test estimates directly the unknown nuisance parameters while invariance principles circumvent the estimation of those parameters by restricting the test to a class of tests. The applicability of these approaches depends on the particular inference problem at hand. In the example given, the GLRT approach is not applicable as there does not appear to be any reasonable way to estimate the unknown noise variance given only quantized output from a limited number of sensors. On the other hand, the invariance approach depends on the ability to obtain a maximum invariant statistic which in many cases may not be possible due to the lack of symmetry of the inference problem. We should also mention that we do not consider Neyman-Pearson criterion in this paper. Indeed, the problem of distributed detection using Neyman-Pearson criterion gets quite complicated expecially when dependent observations are involved [18].

An obvious disadvantage of the proposed method, compared with previous approaches is its high computational complexity. While the implementation of the Gibbs sampler is conceptually straightforward once the model is well understood, the inference procedure does require significantly more computations than, for example, the classical likelihood based fusion rule. For example, most of the numerical examples (with $50,000$ Monte Carlo runs) in Section 3.3 require $2-5$ hours simulation for the Bayesian sampling approach in a stand alone Pentium PC, while for the likelihood based fusion they usually take a few minutes. However, we should mention that MCMC is itself a major breakthrough in statistics that has significantly lowered the computational complexity as compared with the direct sampling method. For inference problems that can be described using hierarchical models, such as the problem we are dealing with here, computational complexity is usually manageable. In fact, most computations are devoted to the 'sampling' process, i.e., generating random samples that follow some specified distribution. Because of the nice property of hierarchical modes, the sampling distributions can often be reduced to the standard known distribution forms of lower dimensionality. For those distribution, efficient sampling techniques have been well developed and documented in, among others, [19,20]. Issues regarding the computational efficiency and the convergence of the Gibbs sampler for the proposed algorithm are currently under investigation.

Chapter 4

Adaptive CFAR Detection For Clutter-edge Heterogeneity Using Bayesian Inference

Radar CFAR detection is addressed in this chapter. Motivated by the frequently encountered problem of clutter-edge heterogeneity, we model the secondary data as a probability mixture and impose a hierarchical model for the inference problem. A two-stage CFAR detector stucture is proposed. Empirical Bayesian inference is adopted in the first stage for training data selection followed by a CFAR processor using the identified homogeneous training set for target detection. One of the advantages of the proposed algorithm is its inherent adaptivity; i.e., the threshold setting is much less sensitive to the nonstationary environment compared with other standard CFAR procedures.

4.1 Introduction

Reliable radar CFAR detection is critical in dynamic clutter and jamming scenarios [21]. The existence of heterogeneities in practical operational environments renders the conventional cell averaging CFAR (CA-CFAR) ineffective. Heterogeneities arise due to the presence of multiple targets and clutter edges [22]. Alternative schemes have been developed to address this issue, including order statistic CFAR (OS-CFAR) and its variations [23–25] as well as various windowing techniques aimed to exclude heterogeneous regions. Nonetheless, each scheme is targeted toward a particular clutter/interfering target scenario. For example, a variation of CA-CFAR, called the greatest of CFAR (GO-CFAR), calculates the average of the leading and lagging windows, respectively, and selects the greater of the two as an estimate of the clutter strength. Clearly the underlying assumption is that the non-homogeneity (e.g., clutter edge) appears in either the leading or lagging window, but not both. Systematic analysis of various CFAR detectors in non-homogeneous background is given in [26]. Attempts have also been made to intelligently select a CFAR scheme based on some homogeneity test statistics [27].

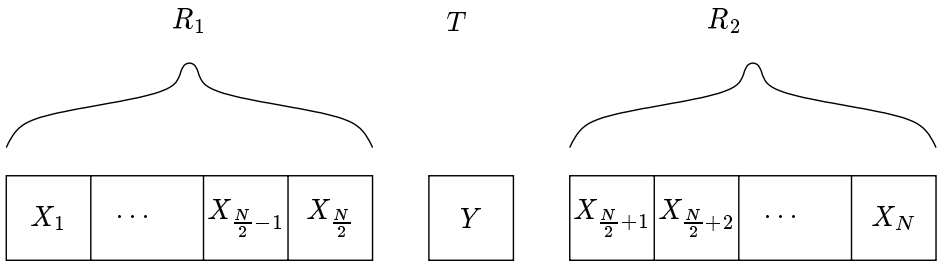


Figure 4.1: Illustration of a CFAR detection problem. R_1 and R_2 are leading and lagging window reference cells and T is the test cell.

Fig. 4.1 is a simple illustration of the CFAR problem under consideration. Our task is to decide if there is a target present in the test cell. For convenience, we assume that the reference cells (also called secondary data) are equally split on either side of the test cell. Here X_n's and Y are the outputs of square law devices that process the in-phase (I) and quadrature (Q) data of the secondary and test cells, respectively. If the background is indeed homogeneous, a GLRT receiver exists where the noise strength is calculated by simply averaging over all reference cells and a threshold for Y can be set that satisfies a certain false alarm constraint. In the presence of a heterogeneity, most CFAR schemes circumvent the direct estimation of the noise statistics. Instead, various non-parametric schemes are adopted for robust detection performance [21].

In this chapter, we propose a probabilistic mixture model to account for the nonhomogeneity of the secondary data in conjunction with Bayesian inference for parameter estimation [28]. We adopt a parametric approach in dealing with possible heterogeneities. A distinction between the proposed algorithm and several existing CFAR approaches is that the former is implemented in two stages. The first stage is a homogeneous region identification procedure, which is then followed by a standard CFAR method such as CA-CFAR applied to the selected homogeneous regions. Notice that reference [27] also adopts a similar approach, though substantial differences exist. In [27], a homogeneity test statistic was applied to predetermined data windows (namely, leading, lagging, and the full window). The data window that appears to be the most homogeneous is selected for clutter statistic estimation. In this chapter, homogeneous regions are identified adaptively and are not limited to any predetermined data window scenarios as in [27].

The proposed approach is particularly suitable for the case of heterogeneities due to the presence of clutter edges. To illustrate the idea, consider the following simple case. Our reference cells consist of two groups corresponding to two different regions with clutter powers μ_0 and μ_1, where $\mu_0 < \mu_1$, and which are otherwise homogeneous within their occupied region. Thus, we have a clutter edge within the reference cells (see Fig. 4.2 for examples). The distribution properties derived from the square law processing of two independent Gaussian processes (in-phase and quadrature components) result in a Rayleigh envelope. For Rayleigh background, we have the following scenario: Those cells associated with the lower level clutter region have observations that follow an exponential distribution with parameter μ_0, while the others follow an exponential distribution with parameter μ_1. Notice that the test cell may belong to either one of the two regions. This

mixture model is deterministic in the sense that every reference cell belongs to one of the two classes. To take advantage of many existing inference tools, we can convert this to a probabilistic mixture model where each cell is associated with one of the two families with a certain probability. Under this framework various inference tools can be used to estimate the mixture model parameters. Candidates include the expectation-maximization (EM) algorithm [29, 30] and various Bayesian inference approaches [13].

In this work, we explore the application of empirical Bayesian inference for parameter estimation. Specifically, a hierarchical model is proposed for the characterization of the non-homogeneous reference cells and the associated priors are estimated using the data rather than chosen *a priori*. A maximum likelihood (ML) estimation algorithm for the unknown parameters has been proposed in [30] where an EM algorithm was developed to iteratively solve the ML estimation problem. However, the inference goal here differs from that of [30]. In [30], the goal is to make a binary decision on whether we have identically distributed exponential random variables, or a mixture of two different exponential random variables. In the current CFAR problem, the objective is to identify the homogeneous clutter region, along with the estimates for the clutter statistics assuming an exponential mixture model. Spatial continuity of each homogeneous clutter region in the presence of a clutter edge will be used to assist the identification process. The estimated clutter statistics will be used for the ensuing CFAR detection.

4.2 CFAR Detection Using Bayesian Inference

4.2.1 Hierarchical modeling of clutter edge heterogeneity

In reference to Fig. 4.1, our goal is to determine whether or not there is a target that dwells in the test cell. Under the Rayleigh clutter assumption, it can be easily established that the likelihood ratio test amounts to the simple thresholding of Y:

$$Y \underset{K}{\overset{H}{\lessgtr}} \tau$$

where H is the target absent hypothesis and K is the target present hypothesis. The choice of τ affects the false alarm level and for CFAR detection, it is desirable to choose τ such that the false alarm probability is maintained at a constant level. Assume that in the target absent case the observation follows an $exp(\mu)$ distribution, then a simple choice is to make $\tau = k\mu$ where $k = -\ln P_{fa}$ to achieve the desired false alarm probability value P_{fa}. In practice, however, μ is usually unknown and needs to be estimated, possibly from the observations in the reference cells. For example, if we assume a homogeneous background, the observations in all reference cells follow the same distribution $exp(\mu)$, so that the CA-CFAR is clearly the optimal approach. Estimating μ is much more complicated in the presence of nonhomogeneity (e.g., a clutter edge) and extensions based on some heuristics have been developed that aim to utilize those reference cells that are considered homogeneous. These include greatest-of CFAR (GO-CFAR) and smallest-of

CFAR (SO-CFAR) and censored cell average CFAR (CCA-CFAR) [31, 32]. For example, in GO-CFAR [33, 34], the average power of leading (R_1) and lagging (R_2) windows are computed and the larger one is used to estimate the clutter strength in the test cell. Another important CFAR scheme is the order statistic based CFAR (OS-CFAR). Rather than averaging over (part of) the reference window, a specific order statistic is chosen as an estimate of the background strength. The order statistic is less sensitive to outliers and OS-CFAR tends to give more robust performance compared with other CFAR detectors. However, OS-CFAR suffers large CFAR loss in terms of signal to clutter power ratio if the background is indeed homogeneous.

In the presence of a clutter edge, the reference cells can be approximated as a mixture of two groups: Those corresponding to either the lower or higher intensity clutter region. With Rayleigh clutter, the two groups of clutter regions have power variation according to exponential distributions such that

$$f(x_n|Z_n = z_n) = (1 - z_n)\frac{1}{\mu_0}e^{-x_n/\mu_0} + z_n\frac{1}{\mu_1}e^{-x_n/\mu_1} \tag{4.1}$$

where $Z_n = 0, 1$ indicates whether cell n belongs to the clutter region with mean value μ_0 or μ_1. The partitioning is deterministic; i.e., each cell belongs to either one of the two families. This deterministic partitioning is not convenient for inference purposes. Therefore, we adopt a probabilistic mixture model. Specifically, we assume that each Z_n is a Bernoulli random variable with success probability θ; i.e., $P[Z_n = 1] = \theta$ and $P[Z_n = 0] = 1 - \theta$. For example, if there is only one clutter type present, ideally we should have all $Z_n \equiv 0$, or equivalently, $\theta = 0$. The prior probability θ is usually unknown. Under the strict Bayesian inference framework, we need to further assign a prior probability for θ. For example, we may assume that θ is uniformly distributed between 0 and 1, or adopt a more general prior in the form of a *Beta* distribution. In this chapter, we use the empirical Bayesian inference procedure: The prior is estimated from the observations. Empirical Bayesian inference provides a compromise between classical inference and Bayesian inference. While the Bayesian inference tools can be utilized in the inference problem, we avoid the choice of prior to prevent any bias introduced by the prior toward the inference goal. Empirical Bayesian is more suitable to scenarios where the inference result is determined to a large extent by the observations rather than the prior. Clearly, for the CFAR problem, when the number of reference cells is moderate to large, empirical Bayesian appears to be a good choice.

The above model is summarized in Fig. 4.3. All observations, X_n, $n = 1, 2, \cdots, N$, follow an exponential mixture model as in (4.1). The Bernoulli variable Z_n's have success probability θ that is common for Z_n. Our goal is to infer the noise/clutter statistics μ_0 and μ_1 as well as the posterior probability that each cell belongs to one of the exponential distributions, i.e., $W_n = P(Z_n = 1|\theta, \mathbf{X})$. Notice that θ serves as a prior, yet it is to be estimated from the data. The estimate of W_n is necessary as it gives an indication as to which region the test cell belongs. Further, it can also be used to improve the CFAR detection performance by utilizing the spatial continuity of different clutter regions as we shall see later.

4.2.2 Training Data Selection Using Bayesian Inference

Given the hierarchical model specification in the previous section, we now estimate the clutter statistics μ_0 and μ_1, as well as θ, the prior probability of Z_n being 1. The posterior probability that $Z_n = 1$, W_n, as it turns out, can be derived directly from the EM algorithm.

The maximum likelihood estimation for μ_0, μ_1, and θ aims to maximize

$$f_{\mu_0,\mu_1,\theta}(\mathbf{X}) = \prod_{i=1}^{N} \left[(1-\theta)\frac{1}{\mu_0} e^{-X_i/\mu_0} + \theta \frac{1}{\mu_1} e^{-X_i/\mu_1} \right]$$

In [30], an expectation-maximization (EM) algorithm was developed for the maximum likelihood estimation of the unknown parameters θ, μ_0, and μ_1. Here, we briefly summarize the iterative procedure.

1. E-step

$$\begin{aligned} \omega_1(i) &\triangleq P(z_i = 1|\theta, X_i) = \frac{\theta \frac{1}{\mu_1} e^{-X_i/\mu_1}}{\theta \frac{1}{\mu_1} e^{-X_i/\mu_1} + (1-\theta)\frac{1}{\mu_0} e^{-X_i/\mu_0}} \\ \omega_0(i) &\triangleq P(z_i = 0|\theta, X_i) = \frac{(1-\theta)\frac{1}{\mu_0} e^{-X_i/\mu_0}}{\theta \frac{1}{\mu_1} e^{-X_i/\mu_1} + (1-\theta)\frac{1}{\mu_0} e^{-X_i/\mu_0}} \end{aligned}$$

2. M-step

$$\begin{aligned} \mu_0 &= \frac{\sum_i \omega_0(i) X_i}{\sum_i \omega_0(i)} \\ \mu_1 &= \frac{\sum_i \omega_1(i) X_i}{\sum_i \omega_1(i)} \end{aligned}$$

and

$$\theta = \frac{1}{N} \sum_{i=1}^{N} \omega_1(i)$$

The EM algorithm involves iteration between the E and M steps until convergence occurs. The estimation for the posterior probability W_n is essentially the same as for $\omega_1(n)$. Hence, it can be derived directly from the EM algorithm; i.e.,

$$W_n = \frac{\theta \frac{1}{\mu_1} e^{-X_i/\mu_1}}{\theta \frac{1}{\mu_1} e^{-X_i/\mu_1} + (1-\theta)\frac{1}{\mu_0} e^{-X_i/\mu_0}}$$

A simple simulation is conducted to determine the effectiveness of the EM algorithm in estimating the parameters. In this example, N is chosen to be 32 and we assume that the first 8 samples correspond to clutter background with mean strength $\mu_1 = 15$ while the rest of the cells follow an exponential distribution with $\mu_0 = 1$ (see Fig. 4.2(a) for an illustration). The EM algorithm yields estimates for the mean values $\hat{\mu_0} = 0.8886$ and $\hat{\mu_1} = 13.0787$. Further, we plot the estimates of the W_n values versus the reference cell index n in Fig. 4.4. We note that they roughly reflect

the separation of different clutter regions; i.e., W_n's with large values are mostly in the first 8 samples. However, it is also easy to see from the figure that due to the randomness of the data within each region, there will be some cells whose estimated posterior probability W_n does not truly reflect its association. This, however, can be somewhat alleviated by taking advantage of the spatial continuity as presented in the next section.

4.2.3 Homogeneous Region Identification Using Spatial Continuity

An important by-product of the EM algorithm is the posterior probability W_n which indicates the group to which each reference cell belongs. These estimates help determine the clutter region in which the test cell dwells and is important in obtaining robust detection performance. For example, if the test cell is in the region with higher clutter intensity (μ_1) but the threshold is determined using μ_0, we may experience excessive false alarms.

Clearly, once we identify the test cell clutter region location, the estimates of the clutter statistics, μ_0 and μ_1, that are obtained from the EM algorithm can be used directly in the CFAR detector. An alternative approach is to re-estimate the clutter statistics by utilizing the output parameters, W_n, $n = 1, \cdots, N$. Notice that the W_n values are the posterior probability estimates that the nth test cell is associated with one of the two groups. Therefore, quantizing the W_n values using threshold 0.5 has the desired property of having minimum error probability for the reference classification. The clutter strengths μ_0 and μ_1 can be re-estimated using the simple averages of the corresponding group of cells distinguished by thresholding the W_n's. This amounts to converting the probabilistic mixture back to the deterministic mixture that is more consistent with the ground truth.

The above discussion also motivates further improvement by utilizing spatial continuity of the clutter regions. In this chapter, a simple heuristic approach is adopted. The association of one particular cell, namely the n^{th} cell within the reference window, to either one of the two distributions is determined not only by its posterior probability W_n but also by its neighbors; i.e.,

$$W_{n-n_s}, W_{n-n_s+1}, \cdots, W_{n+n_s}$$

where $2n_s+1$ is the window size. A simple majority rule is implemented in this chapter to determine the actual value for W_n. If the majority of W_i's within the window (including W_n) belong to $exp(\mu_1)$ (i.e., they have value greater than 0.5), then we claim cell n also belongs to $exp(\mu_1)$. Otherwise, it is assumed to belong to $exp(\mu_0)$. Other heuristic rules, such as mean posterior probability within the sliding window, can also be proposed. More sophisticated probabilistic models that capture spatial continuity can also be developed. Examples include the spatial hidden Markov model.

The above idea is illustrated in Fig. 4.5. Fig. 4.5(a) is the result of direct quantization of W_n, while Fig. 4.5(b) gives the result after post processing using spatial continuity where $n_s = 2$; i.e., the window size for the majority rule is 5. Clearly, Fig. 4.5(b) gives a more accurate account of the association of each cell. That is, it reflects more faithfully the ground truth where only the first 8 cells belong to the high intensity clutter region. Thus, estimation of the clutter statistics based on

this approach tends to be more accurate. We note that while Fig. 4.5(b) matches precisely with the ground truth, this may not always happen. However, the use of spatial continuity tends to improve the overall clutter region identification performance.

4.2.4 CFAR Detection

The training data selection procedure described in the previous section yields a set of data that are assumed to be homogeneous and are of the same clutter type as the test cell. We emphasize again that profiling of the cells using W_n's along with the spatial continuity property allows us to determine the region in which the test cell lies. Based on this selected training data set, various CFAR procedures can be implemented.

Presumably, if the data set is indeed homogeneous, CA-CFAR clearly is the obvious choice due to its optimality for a homogeneous background. If, however, the training data set selected in the first stage still lacks homogeneity (though it should always appear more homogeneous than the original secondary data), more robust CFAR procedures can be implemented on the selected training data set. The proposed CFAR detection procedure is summarized in Fig. 4.6, where the first stage is to identify the homogeneous region using Bayesian inference, followed by standard CFAR procedures applied only to the homogeneous region obtained from stage 1.

An important advantage of the proposed two stage CFAR processor is its inherent adaptivity, hence enhanced robustness, with regard to the changing environment. This is especially true if the mixture is a good approximation of the real scenario hence the first stage yields a relatively homogeneous group of data. Its robustness results since the first stage training data selection can adaptively determine the homogeneous clutter region. Thus the threshold setting is relatively simple to determine — it amounts to the threshold setting for CA-CFAR under homogeneous background! This, however, is not the case for other CFAR processes. For example, to achieve a desired false alarm rate for OS-CFAR, the threshold is determined by a specific clutter edge scenario and may vary drastically from one case to another. This will be addressed further in the next section using some simulation results.

4.3 Numerical Examples

In this section, we present numerical examples to demonstrate the performance of various CFAR detectors, including the EM-CFAR method proposed here. In particular, we consider algorithm performance in various scenarios with the clutter edge location/duration as the parameters. Primary consideration is given to the achievement of good detection performance with minimal variation in threshold to achieve a specified false alarm level. Throughout the simulations, a non-fluctuating target is used and the SNR is computed using the target power versus the average clutter power of the test cell. Also, the sliding window size for spatial continuity is set at 5 for all cases. We investigate a total of four different scenarios as illustrated in Fig. 4.2. Fig. 4.2(a) is an example where there is a clutter edge in the leading window while Fig. 4.2(b) shows the case of a clutter edge

in the lagging window. The difference between the two is that the cell under test dwells in different clutter regions. Fig. 4.2(c) is an example where heterogeneity appears in the lagging window in the form of a high intensity clutter region of finite length. The last example from Fig. 4.2(d) is the case where the test cell dwells in a high power clutter region that sits in the middle of the reference cells. Throughout the examples, a total of 32 reference cells are used (excluding the test cell). The false alarm probability is fixed at 10^{-4} and we use 10^6 Monte Carlo runs to determine the detection probability for various signal-to-noise ratio (SNR) values. The powers for the two different clutter regions are 1 and 15 respectively, resulting in a power difference of approximately $12dB$. The SNR is calculated using the signal power over the power of the clutter region in which the test cell dwells. In the implementation for OS-CFAR, we choose the 20^{th} order statistic as an estimate of the clutter power.

Detection assessment

- Example 1

  In this example, the first 8 cells are assumed to belong to high power clutter and the test cell belongs to the low power clutter region. The results of the probability of detection as a function of SNR for various CFAR procedures are plotted in Fig. 4.7. Clearly the detection probability of the EM based CFAR is among the leading methods in this case.

- Example 2

  The second example differs from the first in that there is a clutter edge in the lagging window and the test cell now resides in the high power clutter region. The results are obtained in Fig. 4.8. Clearly, in this particular example, the detection probability of the EM based CFAR is only slightly worse than the CA-CFAR while it is superior to all other CFAR schemes.

  The slight advantage of CA-CFAR in this case can be explained as follows. Cell averaging is sensitive to outliers in the sense that extremely large values (even if there are only a few of them) may significantly change the mean value. This is why CA-CFAR suffers in detection performance in Example 1 when the first 8 cells are considered outliers. However, in the second scenario, where the target resides in the high level clutter region, the last 8 cells are considered outliers. Yet, their values are lower bounded by zero and hence the performance for CA-CFAR is fairly good in this case.

- Example 3

  The third example is the presence of a high intensity clutter *region* in the lagging window as illustrated in Fig. 4.2(c). In particular, cells 21 to 28 are assumed to belong to high power clutter. The test cell now resides in the low power clutter region. The probability of detection of the various CFAR schemes is obtained in Fig. 4.9. The detection probability of the EM based CFAR is again the best among all CFAR schemes.

- Example 4

In the last example, high intensity clutter appears in the middle of the secondary cells and the test cell falls in the high power region. Notice here that nonhomogeneity appears in both the leading and lagging windows as in Fig. 4.2(d). The detection performance for EM-CFAR is again fairly robust in this case as shown in Fig. 4.10.

From the above examples, we observe that even though the EM-CFAR does not necessarily yield the best detection performance for all clutter edge scenarios, it tends to compare favorably to the specific CFAR scheme that performs best for each scenario. Thus, it is very robust to the change of the clutter edge position/duration.

Threshold assessment

Perhaps the most noteworthy conclusion from the above results is the inherent adaptivity of the EM-CFAR algorithm. We note that the background clutter not only lacks homogeneity, but is also temporally dynamic. For CFAR schemes such as GO-CFAR, threshold changes are required to maintain a constant false alarm probability. However, for EM-CFAR, the background statistics are estimated using a parametric model that itself adapts inherently to each particular scenario. This greatly alleviates the need for threshold adaptation.

To further understand this, we note that the test for various CFAR schemes, including the EM-CFAR, can be summarized as

$$T = \frac{Y}{\hat{\mu}} \quad \underset{K}{\overset{H}{\lessgtr}} \quad \eta \tag{4.2}$$

where Y is the test cell observation and $\hat{\mu}$ is the estimated clutter power for the test cell, e.g., in CA-CFAR, it is the average over secondary data. We observe in the simulations that the thresholds required for the EM-CFAR to maintain the same false alarm probability have far lower standard deviation than all the other schemes. Table 4.1 shows these results for the desired false alarm probability at $P_{fa} = 10^{-4}$. The normalized standard deviation ($\sigma_\eta/\bar{\eta}$ where η is the threshold) is given in the last column. Clearly, the variation of the threshold for the EM based CFAR detector is the smallest among all of them. Further, we note that for a false alarm probability at 10^{-4}, if indeed the true clutter strength is known, then the desired threshold in (4.2) is $\eta = -\ln P_{fa} = 9.21$. Clearly, the thresholds for EM-CFAR in all cases are close to this nominal value. This facilitates the choice of threshold for the CFAR property for a nonstationary environment. Notice, for example, the CA-CFAR applied to the entire secondary data set requires thresholds that vary drastically to maintain a constant false alarm rate for different heterogeneous scenarios.

4.4 Summary and Conclusions

Training data selection for radar CFAR detection in the presence of clutter edges is addressed in this chapter. An adaptive CFAR detection approach is developed where the heterogeneous secondary data are modeled as a mixture of two different distributions, each with different clutter strengths. Parameter estimation for the clutter statistics is carried out using the empirical Bayesian inference

	Example 1	Example 2	Example 3	Example 4	Normalized Standard Deviation
EM-CFAR	12.2831	10.0528	15.6208	11.9686	0.1854
CA-CFAR	3.0254	13.2838	30.2273	2.8762	1.0421
GO-CFAR	1.9625	10.5917	26.7379	1.8334	1.1391
SO-CFAR	13.0278	24.9282	45.1486	11.8765	0.6503
OS-CFAR	2.8416	7.2997	21.4190	2.7450	1.0286

Table 4.1: The thresholds for all CFAR schemes to maintain false alarm probability at 10^{-4}. While the thresholds for EM-CFAR does not vary from case to case, all other schemes have dramatically different threshold for different examples.

procedure where the priors are estimated using the observations. Spatial continuity of the clutter regions is utilized to improve the training data selection. The homogeneous region where the test cell dwells is determined and used for CFAR detection where various standard CFAR procedures can be subsequently applied. Numerical results show that the proposed method compares favorably to competing CFAR detectors. The proposed CFAR procedure is inherently adaptive and therefore suitable for nonstationary environments. Finally and perhaps most significantly, the threshold setting to maintain a certain false alarm rate is fairly insensitive to the nonstationary environment due to the inherent adaptivity in adjusting to different clutter levels.

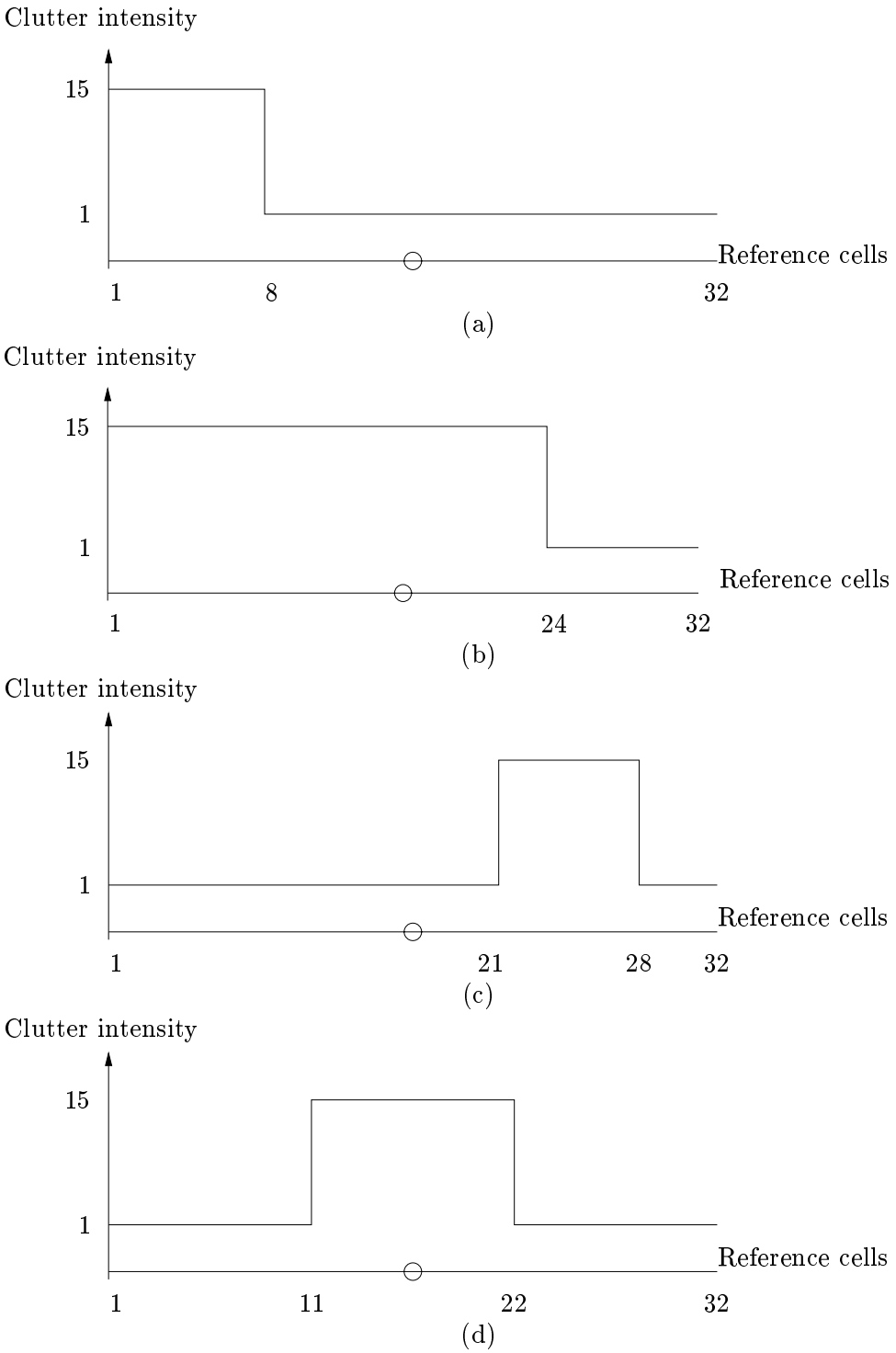


Figure 4.2: Four different clutter edge scenarios. The cell under test is indicated as a circle in the plots.

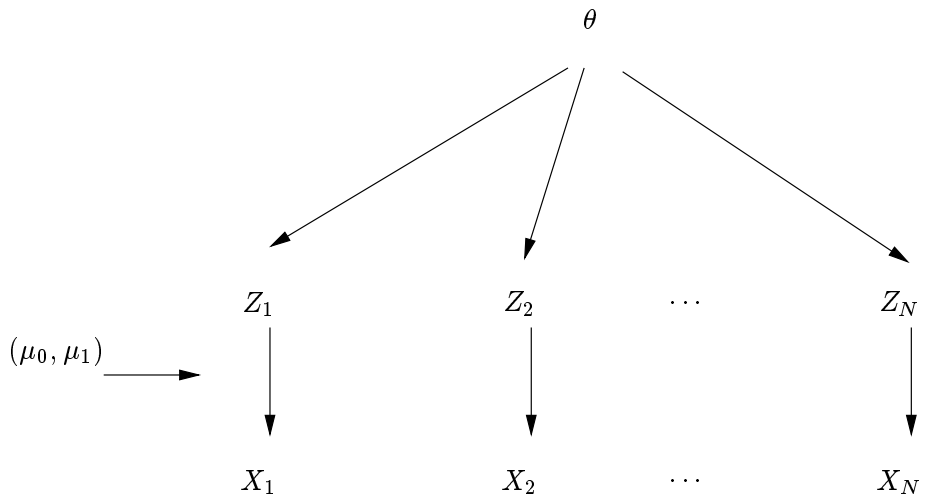


Figure 4.3: Hierarchical modeling for the CFAR parameter estimation problem where the observations are connected through upper layer parameters. Each observation y_n is assumed to be a mixture of two exponential distributions with mean values μ_0 and μ_1. Our goal here is to estimate μ_0 and μ_1, respectively, as well as the posterior probability that each Z_n equals 1, i.e., $P(Z_n = 1|\theta, \mathbf{X})$.

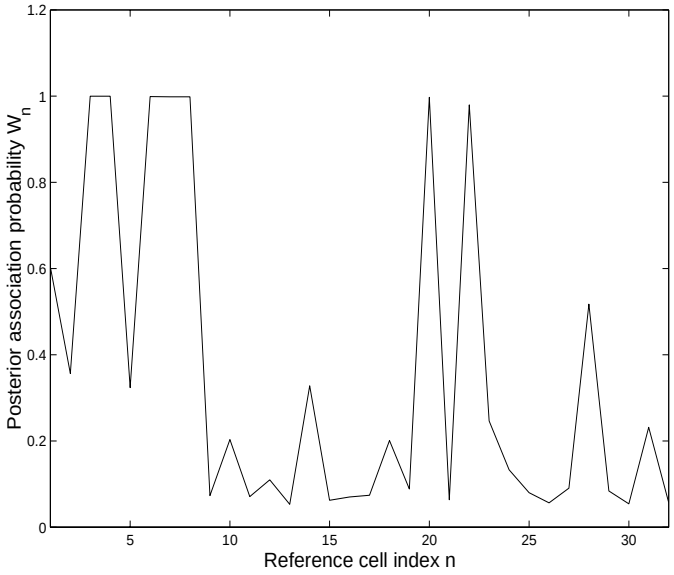
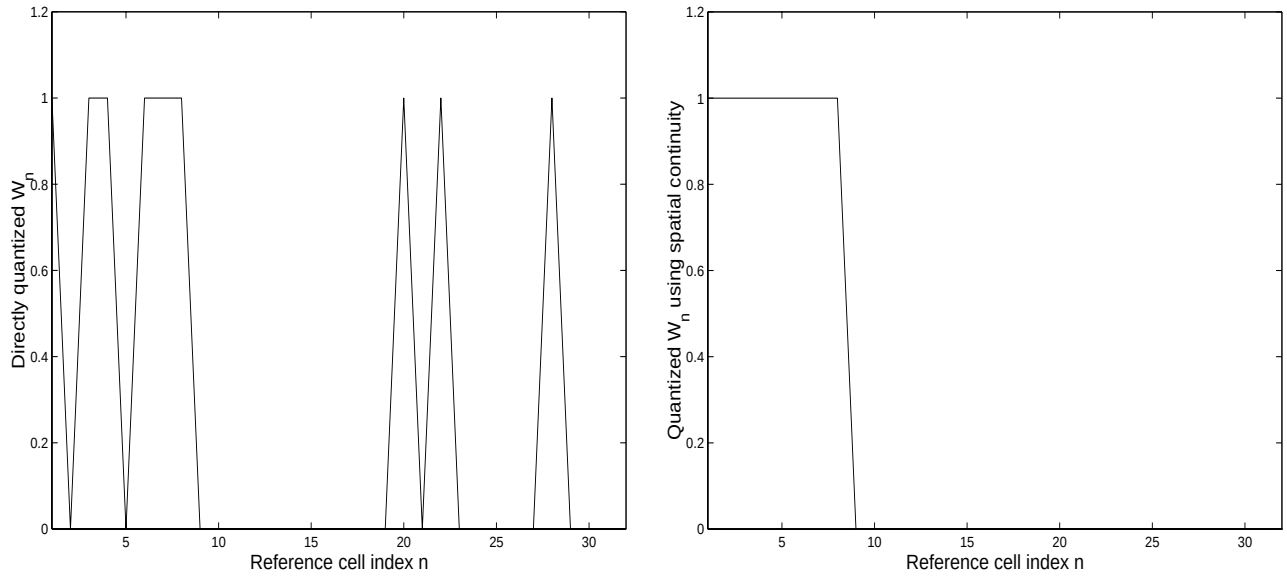


Figure 4.4: Plot of W_n's for the Bernoulli variates in the data model. The ground truth is that the first eight samples follow $exp(\mu_1)$ while the rest of the secondary data follow $exp(\mu_0)$. As expected, the first eight samples have relatively large values as they correspond to the clutter background.



(a) (b)

Figure 4.5: Post processing of the posterior association probability W_n's using the ground truth as used in Fig. 4.4. Figure (a) is the result of direct quantization while (b) utilizes spatial continuity of the clutter region using majority rule with a window size 5.

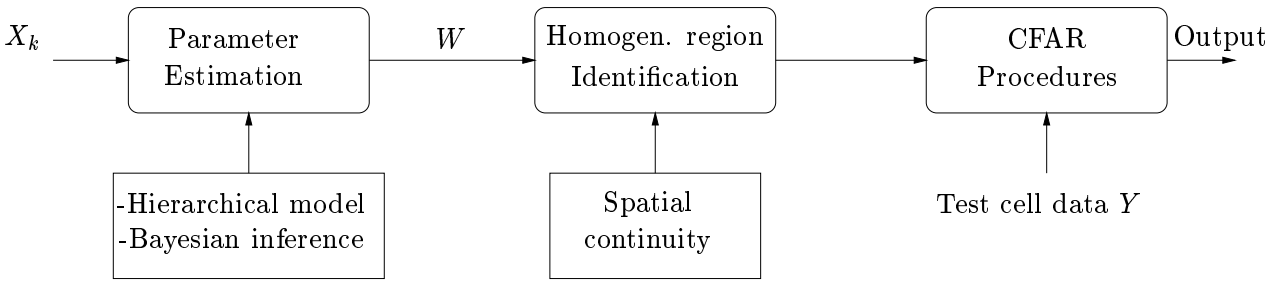


Figure 4.6: Block diagram of the proposed two stage CFAR detector. The first stage (corresponding to the first two blocks) involve homogeneous region identification using Bayesian inference and spatial continuity of clutter regions. The second stage applies standard CFAR procedures to the homogeneous region identified through the estimation procedure.

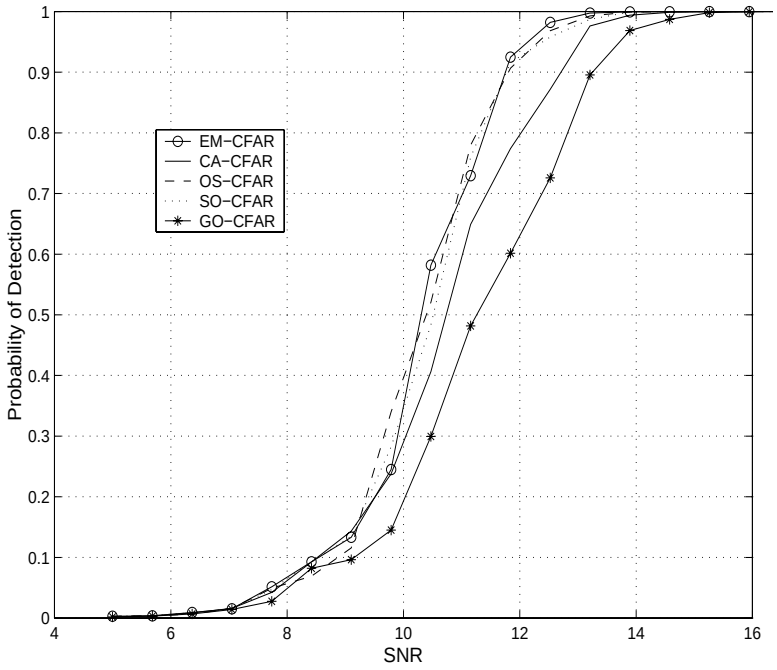


Figure 4.7: The probability of detection as a function of SNR for Example 1 (see Fig. 4.2(a)): $P_{fa} = 10^{-4}$, and 10^6 Monte Carlo runs.

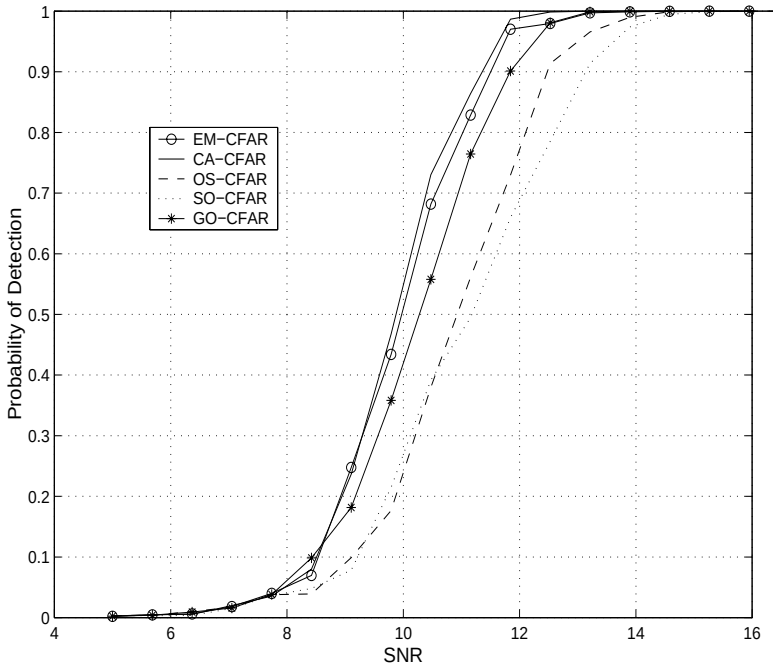


Figure 4.8: The probability of detection as a function of SNR for Example 2 (see Fig. 4.2(b)): $P_{fa} = 10^{-4}$, and 10^6 Monte Carlo runs.

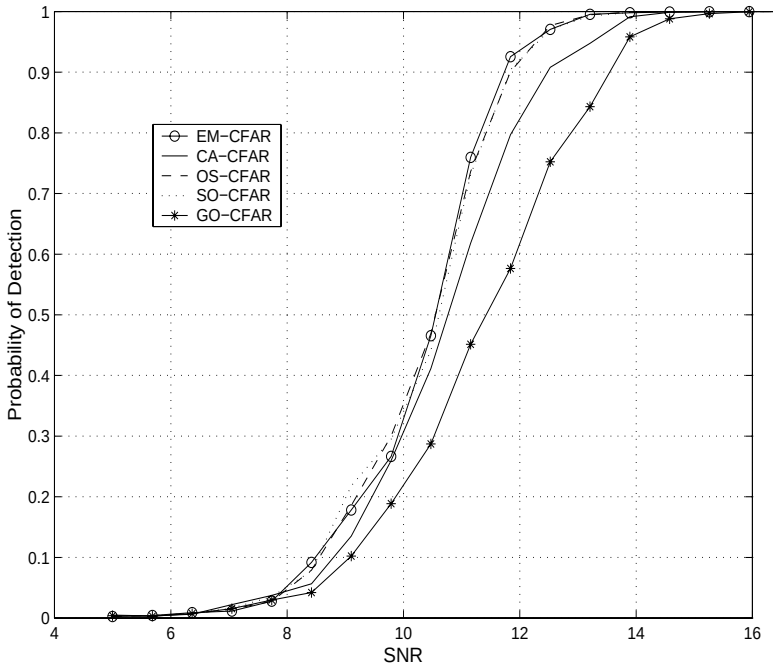


Figure 4.9: The probability of detection as a function of SNR for Example 3 (see Fig. 4.2(c)): $P_{fa} = 10^{-4}$, and 10^6 Monte Carlo runs.

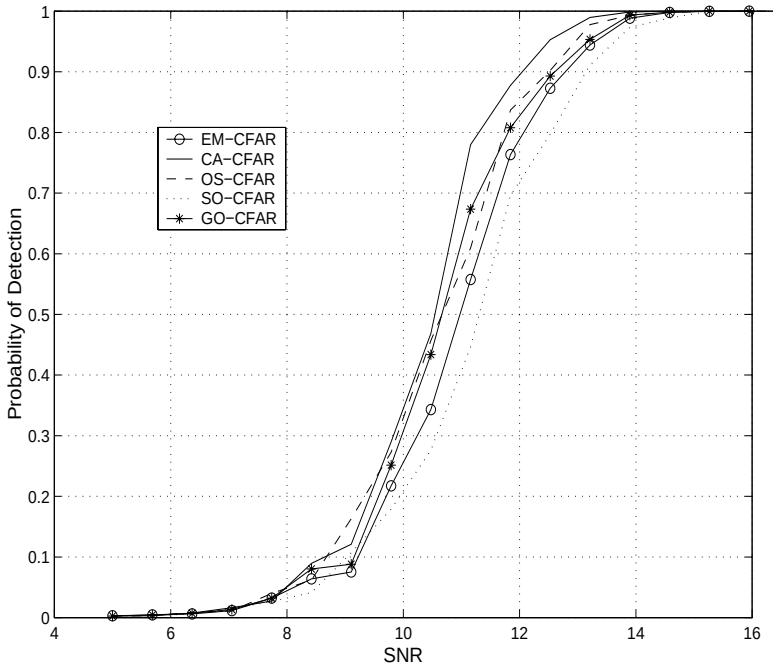


Figure 4.10: The probability of detection as a function of SNR for Example 4 (see Fig. 4.2(d)): $P_{fa} = 10^{-4}$, and 10^6 Monte Carlo runs.

Chapter 5

Clutter Patch Characterization and Identication Using Markov Random Field Models

In this chapter, we address the problem of clutter patch identification based on Markov random field (MRF) models. MRF has long been recognized by the image processing community to be an accurate model to describe a variety of image characteristics such as texture. Here, we use the MRF to model clutter patch characteristics, captured by a radar receiver or radar imagery equipment, due to the fact that clutter patches usually occur in connected regions. Furthermore, we assume that observations inside each clutter patch are homogenous, i.e., observations follow a single probability distribution. We use the Metropolis-Hasting algorithm and the reversible jump Markov chain algorithm to search for solutions based on the Maximum a Posteriori (MAP) criterion. Several examples are provided to illustrate the performance of our algorithm.

5.1 Introduction

Accurate statistical characterization of clutter background is critical to the design of efficient target detection and identification algorithms for radar systems. The conventional model assumes that the return signal consists of a known signal in Gaussian noise. The performance of a radar signal detector based on this conventional model degrades significantly in the presence of a complex clutter background. To improve performance, it is imperative that the clutter background be modeled accurately. One needs to determine the homogenous patches of clutter that occur due to reflections from heterogeneous background. In addition, the underlying probability density functions (PDF) in each clutter patch needs to be identified. Using this information, intelligent detection schemes can be designed that are expected to perform better.

The goal of this paper is to address this important problem and develop an algorithm for clutter patch identification.

Slamani [35] initiated the investigation of the clutter patch identification problem in which the

surveillance volume is divided into homogenous regions based on PDFs of clutter. His methodology is composed of two steps. In the first step, a surveillance volume is separated into clutter patches and background noise regions using an appropriate threshold because noise power in background noise regions is lower than those in clutter patches. An appropriate technique for determining this threshold was also presented. Since clutter patches cannot occur at isolated points, some misclassifications can be removed through a windowing method where the similarity among neighboring pixels is measured and compared. A change to the class at the pixel of interest is made if the number of neighboring pixels conflicting with the pixel of interest exceeds a certain threshold. Clutter patches are divided using the above procedure according to their power levels until they cannot be divided any further. In the second step, each clutter patch is separated into more regions according to their underlying noise distributions. Here, the Ozturk algorithm [36] is used to approximate noise distributions by generating a coordinate from an order statistics of the samples. Then, for a given neighborhood, the coordinate is plotted and distances to the list of possible noise distributions are measured. By using these distances, the border pixels between two adjacent homogenous regions can be determined. Some promising results have been presented in [35]. However, this methodology is intuitive and lacks theoretical justification. Here, we extend the work presented in [35] and develop an algorithm under a statistical framework to automatically identify clutter patches and estimate the underlying noise distributions.

Image segmentation techniques seem to be good candidates for this problem due to similarities between segmentation and clutter patch identification problems. The main objective of both problems is to separate an inhomogeneous region into several homogenous regions according to some features. For image segmentation, these features may be textures or gray levels, whereas, for clutter patch identification, the PDFs are the main concern. In the past few decades, there have been a number of image segmentation algorithms that have been developed for a variety of problems. These fall into two general categories of statistical-based and deterministic algorithms. Under the statistical framework, the Markov random field (MRF) model has received a great deal of attention because a MRF model can characterize the information contained among neighboring pixels quite accurately [37–41]. As a result, we will develop our algorithm based on a MRF model.

5.2 Problem Statement

Let $\mathcal{S}$ be a set of sites (pixels) s, and $\Lambda = \{0, 1, \cdots, L-1\}$ be the phase space (intensity level). Furthermore, let $\mathbf{X}(\mathcal{S}) \in \Lambda^{\mathcal{S}}$ denote a clutter patch (configuration) vector or a clutter patch image (CI). Note that $L \geq 2$ is the number of clutter patches in the scene in which the exact value of L is unknown. Here, we model L as a random variable characterized by the probability mass function $P_L(l)$. We assume that $\mathbf{X}(\mathcal{S})$ satisfies MRF properties with a Gibbs potential $V_C(\mathbf{x})$ [37–40]. Hence, we can write the marginal PDF of $\mathbf{X}(\mathcal{S})$ [40] as

$$\pi_{\mathbf{X}}(\mathbf{x}_i) = \frac{1}{Z_{\mathbf{X}}} exp\left[-\sum_{C \subset \mathcal{S}} V_C(\mathbf{x}_i) \right] \tag{5.1}$$

where $Z_{\mathbf{X}} = \sum_{\mathbf{x}\in\Lambda^S} exp\left[-\sum_{C\subset\mathcal{S}} V_C(\mathbf{x})\right]$ is the normalizing constant. Note that $\mathbf{X}(\mathcal{S})$ is a realization of a clutter scene and $\sum_{C\subset\mathcal{S}} V_C(\mathbf{x})$ is called the Gibbs energy function. These terms will be used extensively in our discussion. Let $\mathbf{Y}(S) \in \mathcal{R}^{\mathcal{S}}$ be the associated clutter or noise vector whose observations in pixels s_i and s_j are statistically independent given the CI, i.e.,

$$p\left(y(s_i), y(s_j)|\mathbf{X}(\mathcal{S})\right) = p\left(y(s_i)|\mathbf{X}(\mathcal{S})\right) p\left(y(s_j)|\mathbf{X}(\mathcal{S})\right) \tag{5.2}$$

where p is the probability density function of clutter or noise. Furthermore, the probability density function (PDF) of clutter or noise at a site depends only on the type of clutter patch and observation at that site, i.e.,

$$p\left(y(s_i)|\mathbf{X}(\mathcal{S})\right) = f\left(y(s_i), x(s_i)\right) \tag{5.3}$$

where f(a,b) denotes a PDF.

In radar and sonar systems, we usually assume that the clutter can be modeled as one of several known distribution types. For example, Slamani [35] assumed that noise can be Rayleigh, Weibull, lognormal or K distributions. Here, we also assume that f(a,b) can only come from a known family of distributions that is

$$f\left(y(s_i), x(s_i)\right) \in \{f_m(y(s_i), \theta_m)\}_{m=\{1,\cdots,M\}} \tag{5.4}$$

where M indicates the size of the family of PDFs, and θ_m is the parameter vector corresponding to the underlying PDF.

We formulate the clutter patch identification problem as an M-ary hypothesis testing problem where each hypothesis corresponds to a different CI. For a given CI (hypothesis), the observation volume is divided into several homogenous regions of clutters. We note again that the term "homogenous" implies that the PDFs of observations at every pixel inside a clutter patch are identical and independent. Furthermore, since we formulate our problem as M-ary hypothesis testing problem, techniques developed to solve signal detection problems can be employed and we provide our methodology in the next section.

5.3 Optimum Clutter Patch Identification Algorithm

The maximum a posteriori (MAP) criterion [2, 42] is used for identifying clutter patches in our work. This criterion is expressed as

$$\mathbf{X}_k = arg\left\{\max_{\mathbf{X}_j}\left[P(\mathbf{X}_j|\mathbf{Y}=\mathbf{y})\right]\right\} \tag{5.5}$$

From Bayes' rule, (5.5) can be rewritten as

$$\mathbf{X}_k = arg\left\{\max_{\mathbf{X}_j}\left[\frac{P(\mathbf{Y}=\mathbf{y})|\mathbf{X}_j)P(\mathbf{X}_j)}{P\mathbf{Y}=\mathbf{y})}\right]\right\} \tag{5.6}$$

Since $P(\mathbf{Y}=\mathbf{y})$ is independent of $\mathbf{X}_k$, the above equation reduces to

$$\mathbf{X}_k = arg\left\{\max_{\mathbf{X}_j}\left[P(\mathbf{Y}=\mathbf{y})|\mathbf{X}_j)P(\mathbf{X}_j)\right]\right\} = arg\left\{\max_{\mathbf{X}_j}\left[\left(\prod_{s\in\mathcal{S}} P(y(s)|x_j(s))\right)P(\mathbf{X}_j)\right]\right\} \tag{5.7}$$

Distribution	PDF
1. T-2 Gumbel	$\gamma y^{-\gamma-1} exp(-y^{-\gamma}) \quad 0 < y < \infty$
2. Gamma	$\frac{1}{\Gamma(\gamma)} exp(-y) y^{\gamma-1} \quad 0 < y < \infty$
3. Pareto	$\frac{\gamma}{y^{\gamma+1}} \quad y > 1$
4. Weibull	$\gamma y^{\gamma+1} exp(-y^{\gamma}) \quad y > 0$
5. Lognormal	$\frac{1}{y\gamma\sqrt{2\pi}} exp\left[-\frac{(log(y)/\gamma)^2}{2}\right] \quad y > 0$
6. K-distribution	$\frac{2}{\Gamma(\gamma)} \left(\frac{y}{2}\right)^{\gamma} K_{\gamma-1}(y) \quad y > 0$
7. Beta	$\frac{1}{B(\gamma,\delta)} y^{\gamma-1}(1-y)^{\delta-1} \quad 0 < y < 1$

Table 5.1: Conventional forms of the PDFs

Substituting (5.1)-(5.3) into (5.7) and recognizing that the number of clutter patches is random, we have

$$\mathbf{X}_k = arg \left\{ \max_{\mathbf{X}_j} \left[\left(\prod_{s \in \mathcal{S}} P_n(y(s)|x_j(s)) \right) \frac{1}{Z_{\mathbf{X}}} exp\left(- \sum_{C \subset \mathcal{S}} V_C(\mathbf{x}_j) \right) P_L(l_j) \right] \right\} \tag{5.8}$$

where l_j is the number of clutter patches in a CI $\mathbf{x}_j$. We note again that $P_L(\cdot)$ is a probability mass function associated with L. Using (5.4), the above equation can be written as

$$\mathbf{X}_k = arg \left\{ \max_{\mathbf{X}_j} \left[\left(\prod_{s \in \mathcal{S}} f_{m(x_j(s))}(y(s), \theta_{m(x_j(s))}) \right) \frac{1}{Z_{\mathbf{X}}} exp\left(- \sum_{C \subset \mathcal{S}} V_C(\mathbf{x}_j) \right) P_L(l_j) \right] \right\} \tag{5.9}$$

where $m(x_l(s))$ denotes the mth type of PDF.

In practice, the direct minimization of equation (5.9) is not feasible due to the enormous number of possible CIs. Moreover, parameter vectors associated with each clutter patch are generally unknown. Therefore, there is a need for a more efficient way to obtain the solution of (5.9) and estimate the unknown parameters. Here, we will employ the Metropolis-Hasting algorithm and the reversible jump Markov chain algorithm [43] together with ML estimation to search for the solution of (5.9) and estimate unknown parameters simultaneously.

5.4 Numerical Results

Here, we choose T-2 Gumbel, Gamma, Pareto, Weibull, Lognormal, K-distribution, and Beta distributions to form the allowed set of PDFs. These PDFs are given in Table 5.1. In this example, the accuracy of our algorithm is illustrated by using the simulated CI displayed in Figure 5.1 whose intensity levels are black, gray and white, respectively. Here, white indicates background noise whereas black and gray indicate clutter patches with different distributions. As mentioned in [35], the background noise region usually has Rayleigh distribution which is equivalent to Weibull distribution with shape parameter 2. For this distribution, we choose the location and scale parameters to be 0.88 and 0.46, respectively.

Figure 5.1: The clutter patch image.

For two clutter patches, we choose a Weibull distribution with location, scale and shape parameters 4.5, 3.0, and 1.5, respectively for the first patch, and a Lognormal distribution with location, scale and shape parameters, 3.6, 4.0 and 0.89, respectively for the other. The resulting observed image is shown in Figure 5.2. Moreover, we assume that the number of clutter patches is a number between two and five each occurring with equal probability. Hence, the maximum number of clutter patches is five.

Next, the observed image is submitted to our proposed algorithm, and the resulting CIs after 0, 50, 100, 200, 300 and 500 iterations are displayed in Figure 5.3 (a)-(f), respectively. Initially, the clutter patches and background noise region have misclassifications in the boundary regions. As the number of iterations increases, the resulting CI approaches the true model in Figure 5.1. After 500 iterations, they are almost identical except for some isolated misclassified pixels. The corresponding results on parameter estimation and distribution approximation are excellent, and they will be provided in the full paper.

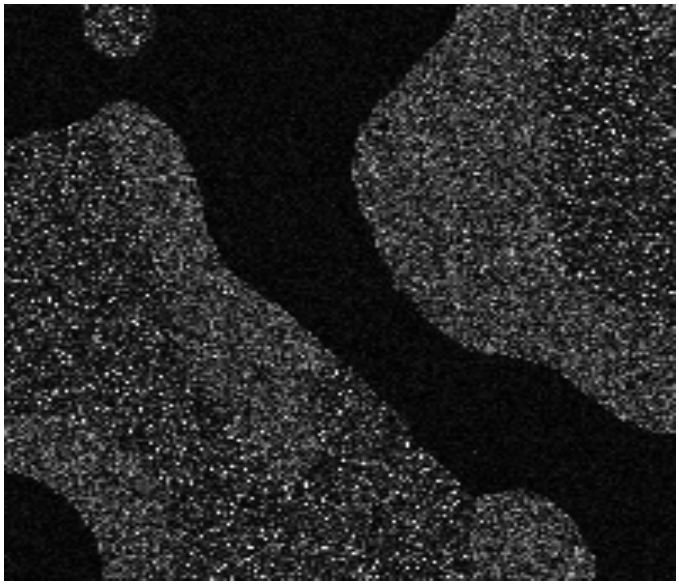


Figure 5.2: An observed image.

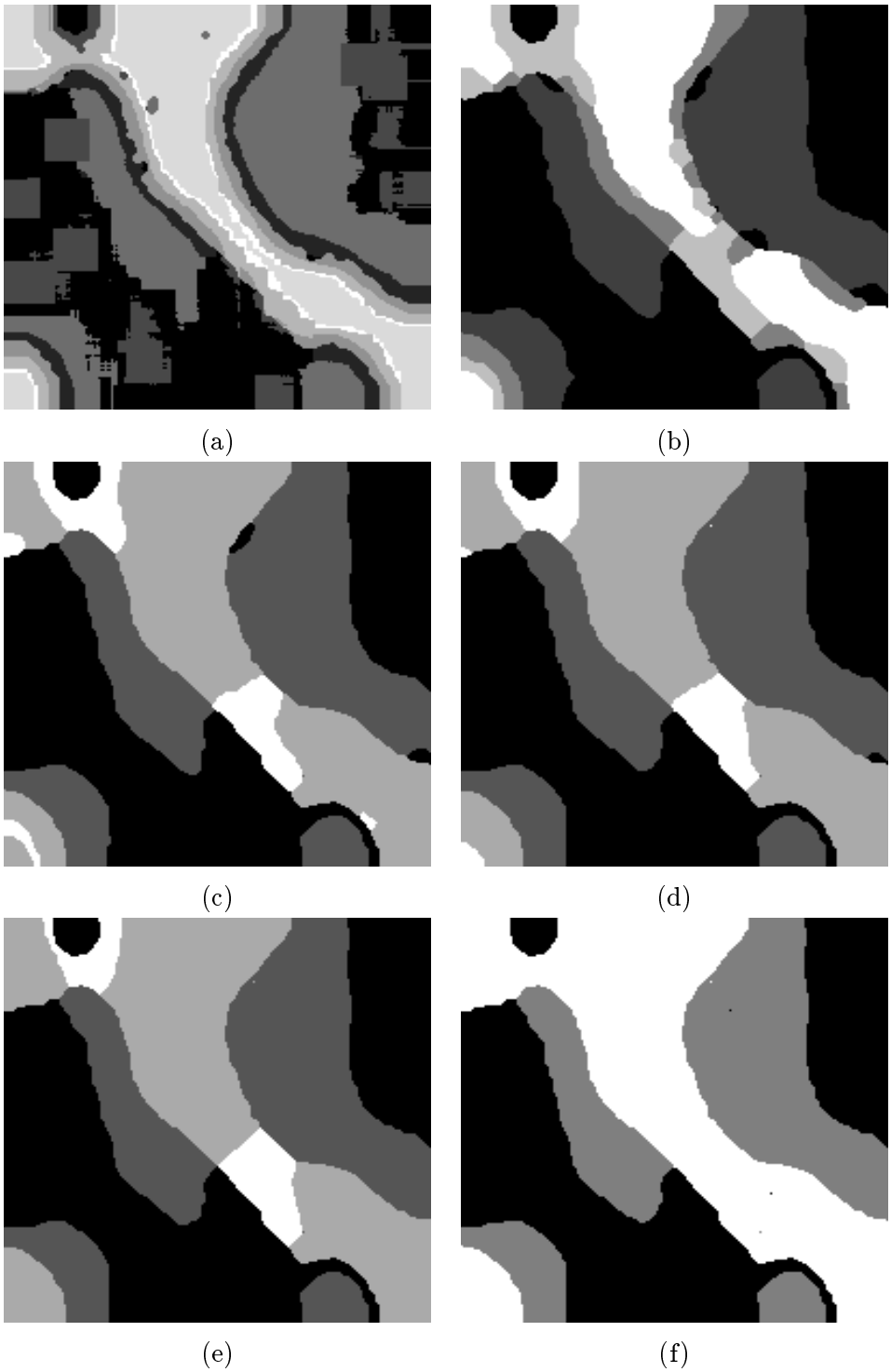


Figure 5.3: : The resulting CI after (a) 0; (b) 50; (c) 100; (d) 200; (e) 300; (f) 500 iterations.

Chapter 6

A GLRT for Multichannel Radar Detection in the Presence of both SIRP Clutter and Additive White Gaussian Noise

Motivated by multi-channel radar detection applications in the presence of both Gaussian and non-Gaussian disturbance, we develop maximum likelihood parameter estimates for spherically invariant random processes (SIRP) in the presence of white Gaussian noise. Both cases with known and unknown white noise variance are treated. As the estimators do not admit closed-form solutions, numerical iterative procedures are developed that are guaranteed to at least converge to the local maximum. The developed estimate allows us to construct a generalized likelihood ratio test (GLRT) for the detection of a signal with constant but unknown amplitude embedded in both Gaussian noise and SIRP disturbances. This new GLRT compares favorably to existing detection schemes that neglect the existence of white Gaussian noise.

6.1 Introduction

Multichannel radar detection considers the detection of the possible presence of a target at a given steering direction in the presence of clutter/noise disturbance. For air-borne high resolution radars operating at low gazing angles, spherically invariant random process (SIRP) has emerged as a viable model to describe the backscattering process. For this clutter model, the clutter vector $\mathbf{c}$ is expressed as $\mathbf{c} = s\mathbf{g}$ where $\mathbf{g}$ is complex Gaussian with covariance matrix $\mathbf{\Sigma}$ and s is a non-negative scalar, unknown random clutter component (also called texture component) statistically independent of $\mathbf{g}$. The power variation of ground clutter among range cells is captured by the variation of s while the Gaussianity is dictated by the central limit theorem applied locally to each

range cell. Thus the multichannel radar detection in the presence of SIRP clutter and additive white Gaussian noise (AWGN) can be formulated as the following hypothesis testing problem

$$\begin{aligned} \mathbf{H}_0 \quad & \mathbf{x} = s\mathbf{g} + \mathbf{n} \\ \mathbf{H}_1 \quad & \mathbf{x} = a\mathbf{v} + s\mathbf{g} + \mathbf{n} \end{aligned} \tag{6.1}$$

where the complex observation data vector $\mathbf{x} \in C^N$ where N is the vector size*; $\mathbf{v}$ is the steering vector, a is the unknown signal amplitude, and $\mathbf{n}$ is the AWGN vector. Examples for SIRP clutter [44–46] include the K distribution and Weibull distribution for specific shape parameter values. It is worth mentioning that SIRP belongs to a widely referenced class of random processes, the so-called compound-Gaussian process when the texture component remains stationary for one coherent processing interval.

While much effort has been undertaken in finding a good detector for signals embedded in SIRP clutter, most existing work assumes a clutter-only model; i.e., the presence of additive white Gaussian noise at the receiver is largely ignored. Consider the clairvoyant case of known $\mathbf{\Sigma}$, i.e., the covariance structure of $\mathbf{g}$ is known. In the absence of white Gaussian noise $\mathbf{n}$, the maximum likelihood (ML) estimate of the unknown parameters, namely the signal amplitude a and the scalar power term for the SIRP component s, can be derived straightforwardly. Substituting the ML estimate under the two hypotheses into the likelihood ratio for the hypothesis testing problem [47], one arrives at the well-known test statistic in the form of

$$\Gamma_1 = \frac{|\mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{v}|^2}{\left(\mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x}\right)\left(\mathbf{v}^H \mathbf{\Sigma}^{-1} \mathbf{v}\right)} \tag{6.2}$$

We remark here that this test statistic has been independently developed in [48] as an asymptotically optimum test for radar detection in SIRP clutter using the representation theorem for SIRP derived in [46]. Since this test statistic added to the matched filter detector a normalizing constant $\mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x}$, we will term it the Normalized Matched Filter (NMF). We note that the test statistic bears the same form as the ACE (adaptive coherence/cosine estimator) test developed for a Gaussian disturbance model with a scale change between test and training data [49].

Disregarding the presence of the AWGN in the detection problem was largely based on the premise that clutter power is usually several magnitudes higher than that of the AWGN, thus making the presence of AWGN seemingly irrelevant. It was pointed out, however, that the presence of AWGN causes the NMF statistic to lose the desired CFAR property (see, e.g., [50]). Here we further demonstrate that, even under extreme power disparity between clutter and noise, a carefully constructed detection statistic can significantly outperform that of (6.2) which neglects the presence of AWGN. To do so, we develop in this chapter a maximum likelihood parameter estimation procedure for SIRP model in the presence of AWGN. Notice that for SIRP, one can develope a Bayesian estimator by utilizing the parametric model imposed on the texture component s. We adopt, however, the ML approach thus treating s in each range cell as an unknown constant,

*In the context of space time processing, $N = JL$ where J is the number of antenna elements and L is the number of pulses within one coherent processing interval.

much in the same way as that of [47]. The estimators can be applied to any SIRP clutter model regardless of the different prior on s. The developed estimates are used to construct a new GLRT that takes into account both SIRP and the white Gaussian disturbances. Throughout this work, we assume that the covariance matrix $\mathbf{\Sigma}$ is known as in [47].

6.2 ML Estimate of Compound-Gaussian Parameters

For airborne radar applications with non-Gaussian clutter, while the clutter texture power (the s parameter) at the test cell is usually unknown, knowledge of the additive white Gaussian noise power, σ^2, may be available from the operational system. We therefore distinguish the following two cases: (1) σ^2 known and (2) σ^2 unknown.

6.2.1 Known σ^2

For the known noise power case, we assume, without loss of generality, that $\sigma^2 = 1$ as the observations can be properly normalized. We need, therefore, to solve the ML equations for 1) s under the $\mathbf{H}_0$ hypothesis; and 2) s and a under the $\mathbf{H}_1$ hypothesis.

ML estimate of s under $\mathbf{H}_0$

Under $\mathbf{H}_0$, the only unknown parameter involved is s and we have the likelihood function:

$$L(s;\mathbf{x}) \propto \frac{1}{\|\mathbf{M}\|}\exp\left(-\mathbf{x}^H\mathbf{M}^{-1}\mathbf{x}\right)$$

Since $\mathbf{\Sigma}$ is assumed to be positive definite and Hermitian, $\mathbf{\Sigma}$ can be diagonalized by a unitary transformation (a.k.a., eigen decomposition)

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$$

where $\mathbf{U}$ is a unitary matrix and $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements, say, λ_i for $i = 1, \cdots, N$, are real positive. Then

$$\mathbf{M} = s\mathbf{\Sigma} + \mathbf{I} = \mathbf{U}(s\mathbf{\Lambda} + \mathbf{I})\mathbf{U}^H$$

From this, we get

$$\begin{aligned}
\|\mathbf{M}\| &= \prod_{i=1}^{N}(s\lambda_i + 1) \\
\mathbf{M}^{-1} &= \mathbf{U}\left(diag\left(\frac{1}{s\lambda_1+1}, \cdots, \frac{1}{s\lambda_N+1}\right)\right)\mathbf{U}^H
\end{aligned}$$

Therefore

$$\frac{\partial\|\mathbf{M}\|}{\partial s} = \sum_{i=1}^{N}\lambda_i\prod_{j\neq i}(s\lambda_j+1) = Tr(\mathbf{\Sigma}\mathbf{M}^{-1})\|\mathbf{M}\| \tag{6.3}$$

$$\frac{\partial\mathbf{M}^{-1}}{\partial s} = \mathbf{U}\left(diag\left(\frac{-\lambda_1}{(s\lambda_1+1)^2}, \cdots, \frac{-\lambda_N}{(s\lambda_N+1)^2}\right)\right)\mathbf{U}^H = -\mathbf{\Sigma}\mathbf{M}^{-2} \tag{6.4}$$

where $Tr(\mathbf{A})$ is the trace of matrix $\mathbf{A}$. The last equality follows from the unitary property of $\mathbf{U}$. By taking the derivative of $L(s;\mathbf{x})$ with respect to s and setting it to 0, we get

$$\frac{\partial L(s;\mathbf{x})}{\partial s} = -\frac{1}{\|\mathbf{M}\|^2}\frac{\partial \mathbf{M}^{-1}}{\partial s}\exp\left(-\mathbf{x}^H\mathbf{M}^{-1}\mathbf{x}\right) - \frac{1}{\|\mathbf{M}\|}\exp\left(-\mathbf{x}^H\mathbf{M}^{-1}\mathbf{x}\right)\mathbf{x}^H\frac{\partial \mathbf{M}^{-1}}{\partial s}\mathbf{x} = 0$$

From (6.3) and (6.4), we have

$$Tr(\mathbf{\Sigma}\mathbf{M}^{-1}) = \mathbf{x}^H\mathbf{\Sigma}\mathbf{M}^{-2}\mathbf{x} \tag{6.5}$$

ML estimates of a and s under $\mathbf{H}_1$

Under $\mathbf{H}_1$, both a and s are unknown and the likelihood function is

$$L(a,s;\mathbf{x}) \propto \frac{1}{\|\mathbf{M}\|}\exp\left(-(\mathbf{x}-a\mathbf{v})^H\mathbf{M}^{-1}(\mathbf{x}-a\mathbf{v})\right)$$

Taking the derivative with respect to a and setting it equal to zero, we have

$$a = \frac{\mathbf{v}^H\mathbf{M}^{-1}\mathbf{x}}{\mathbf{v}^H\mathbf{M}^{-1}\mathbf{v}} \tag{6.6}$$

The ML equation regarding s can be derived in a very similar fashion as that under $\mathbf{H}_0$ hypothesis and we get

$$Tr(\mathbf{\Sigma}\mathbf{M}^{-1}) = (\mathbf{x}-a\mathbf{v})^H\mathbf{\Sigma}\mathbf{M}^{-2}(\mathbf{x}-a\mathbf{v}) \tag{6.7}$$

Thus the ML estimates for a and s are the solutions to the two nonlinear equations (6.6) and (6.7).

6.2.2 Unknown σ^2

If σ^2 is unknown, we also need to find its ML estimate under $\mathbf{H}_0$ and $\mathbf{H}_1$. Define the covariance matrix $\mathbf{M} = s\mathbf{\Sigma} + \sigma^2\mathbf{I}$, the new estimates are developed below.

ML estimates of s and σ^2 under $\mathbf{H}_0$

Under $\mathbf{H}_0$, we have the likelihood function:

$$L(s,\sigma^2;\mathbf{x}) \propto \frac{1}{\|\mathbf{M}\|}\exp\left(-\mathbf{x}^H\mathbf{M}^{-1}\mathbf{x}\right)$$

Since

$$\mathbf{M} = s\mathbf{\Sigma} + \sigma^2\mathbf{I} = \mathbf{U}(diag(s\lambda_1+\sigma^2,\cdots,s\lambda_N+\sigma^2))\mathbf{U}^H$$

We get

$$\begin{aligned}\|\mathbf{M}\| &= \prod_{i=1}^{N}(s\lambda_i+\sigma^2)\\ \mathbf{M}^{-1} &= \mathbf{U}\left(diag\left(\frac{1}{s\lambda_1+\sigma^2},\cdots,\frac{1}{s\lambda_N+\sigma^2}\right)\right)\mathbf{U}^H\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\partial\|\mathbf{M}\|}{\partial\sigma^2} &= \sum_{i=1}^{N}\frac{\|\mathbf{M}\|}{s\lambda_i+\sigma^2} = Tr(\mathbf{M}^{-1})\|\mathbf{M}\|\\ \frac{\partial\mathbf{M}^{-1}}{\partial\sigma^2} &= \mathbf{U}\left(diag\left(\frac{1}{(s\lambda_1+\sigma^2)^2},\cdots,\frac{1}{(s\lambda_N+\sigma^2)^2}\right)\right)\mathbf{U}^H = \mathbf{M}^{-2}\end{aligned}$$

Taking the derivative of $L(s, \sigma^2; \mathbf{x})$ with respect to s and σ^2 and setting them to zero, we obstain

$$\left.\begin{array}{rcl} Tr(\mathbf{\Sigma M}^{-1}) & = & \mathbf{x}^H \mathbf{\Sigma M}^{-2} \mathbf{x} \\ Tr(\mathbf{M}^{-1}) & = & \mathbf{x}^H \mathbf{M}^{-2} \mathbf{x} \end{array}\right\} \tag{6.8}$$

ML estimates of a, s and σ^2 under $\mathbf{H}_1$

The likelihood function becomes

$$L(a, s, \sigma^2; \mathbf{x}) \propto \frac{1}{\|\mathbf{M}\|} \exp\left(-(\mathbf{x} - a\mathbf{v})^H \mathbf{M}^{-1} (\mathbf{x} - a\mathbf{v})\right)$$

and by taking the derivative of $L(a, s, \sigma^2; \mathbf{x})$ with respect to s, σ^2 and a and setting them to zero, we get

$$\left.\begin{array}{rcl} a & = & \frac{\mathbf{v}^H \mathbf{M}^{-1} \mathbf{x}}{\mathbf{v}^H \mathbf{M}^{-1} \mathbf{v}} \\ Tr(\mathbf{\Sigma M}^{-1}) & = & (\mathbf{x} - a\mathbf{v})^H \mathbf{\Sigma M}^{-2} (\mathbf{x} - a\mathbf{v}) \\ Tr(\mathbf{M}^{-1}) & = & (\mathbf{x} - a\mathbf{v})^H \mathbf{M}^{-2} (\mathbf{x} - a\mathbf{v}) \end{array}\right\} \tag{6.9}$$

6.3 Numerical procedure for solving the ML equations

The set of nonlinear equations developed in Section 6.2 for solving the maximum likelihood estimates do not admit closed-form solutions even for the simplest possible case, namely the estimate of s under $\mathbf{H}_0$ with σ^2 known (equation (6.5)). Numerical procedures are now considered to solve these set of equations. While many standard numerical methods such as the Newton method [51] can be applied to solve these equations, we found that a simple bisection algorithm works well in obtaining reasonably good results fairly efficiently, especially when the clutter to noise power ratio (CNR) is large. Indeed, for large CNR, the solutions to the nonlinear equations are almost always unique, which makes the bisection algorithm an appealing candidate due to its simplicity in terms of implementation.

We use the simple example of (6.5) to illustrate the implementation of the bisection algorithm. Rewrite (6.5) as

$$f(s) \stackrel{\triangle}{=} Tr(\mathbf{\Sigma M}^{-1}) - \mathbf{x}^H \mathbf{\Sigma M}^{-2} \mathbf{x} = 0 \tag{6.10}$$

where under $\mathbf{H}_0$, $\mathbf{x} \sim CN(0, \mathbf{M})$. We use the following bisection method to obtain the solution of $f(s) = 0$.

1. Find $s_0^- < s_0^+$ such that $f(s_0^-) < 0 < f(s_0^+)$. Set $k = 0$.
2. If $|s_k^+ - s_k^-| < \epsilon$ for a given tolerance ϵ, then $s_{final} = \frac{1}{2}(s_k^+ + s_k^-)$;
3. Else, $s_{k+1} = \frac{1}{2}(s_k^+ + s_k^-)$; and
   - If $f(s_{k+1}) = 0$, then $s_{final} = s_{k+1}$;
   - Else If $f(s_{k+1}) < 0$, then $s_{k+1}^- = s_{k+1}$ and $s_{k+1}^+ = s_k^+$;
   - Else If $f(s_{k+1}) > 0$, then $s_{k+1}^+ = s_{k+1}$ and $s_{k+1}^- = s_k^-$.

4. $k = k + 1$. Go to 2.

The bisection algorithms for other sets of nonlinear equations are slightly more complicated. For example, with known σ^2 and under $\mathbf{H}_1$, we need also estimate a in addition to s. However, since a has a closed-form solution given s, one only needs to insert a step at each iteration for a.

Next, we discuss the existence and uniqueness of the solutions to the nonlinear equations. We show that under the large CNR, there always exist solutions to the set of nonlinear equations. Consider the case for hypothesis $\mathbf{H}_0$ with known noise variance, therefore we only have a single variable to deal with.

For $f(s)$ as in equation (6.10), we know that $\mathbf{\Sigma}$ can be diagonalized by a unitary matrix $\mathbf{U}$, i.e.,

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$$

where $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_N)$. Define $\mathbf{y} = \mathbf{U}^H\mathbf{x}$, so that

$$\begin{aligned} f(s) &= Tr(\mathbf{\Sigma}\mathbf{M}^{-1}) - \mathbf{x}^H\mathbf{\Sigma}\mathbf{M}^{-2}\mathbf{x} \\ &= Tr\left(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\left(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H + \sigma^2\mathbf{I}\right)^{-1}\right) - Tr\left(\mathbf{x}^H\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\left(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H + \sigma^2\mathbf{I}\right)^{-2}\mathbf{x}\right) \\ &= \sum_{i=1}^{N}\frac{\lambda_i}{s\lambda_i + \sigma^2} - \sum_{i=1}^{N}\frac{|y_i|^2\lambda_i}{s\lambda_i + \sigma^2} \\ &= \sum_{i=1}^{N}\frac{\lambda_i}{s\lambda_i + \sigma^2}\left(1 - \frac{|y_i|^2}{s\lambda_i + \sigma^2}\right) \end{aligned} \tag{6.11}$$

where we define $\mathbf{y} = \mathbf{U}^H\mathbf{x}$ hence the covariance matrix for $\mathbf{y}$ is $s_0\mathbf{\Lambda} + \sigma^2\mathbf{I}$ and s_0 is assumed to be the true underlying scale parameter.

Define

$$\begin{aligned} g_i(s) &= 1 - \frac{|y_i|^2}{s\lambda_i + \sigma^2} \\ f_i(s) &= \frac{\lambda_i}{s\lambda_i + \sigma^2}g_i(s) \end{aligned}$$

Then $g_i(s)$ is monotonically increasing in s with

$$\begin{aligned} \lim_{s\to\infty} g_i(s) &= 1 \\ \lim_{s\to 0} g_i(s) &= 1 - \frac{|y_i|^2}{\sigma^2} \end{aligned}$$

Define P_0 as the probability that the limit at zero for $g_i(s)$ is less than zero, i.e.,

$$\begin{aligned} P_0 &= P\left(1 - \frac{|y_i|^2}{\sigma^2} < 0\right) \\ &= P(|y_i|^2 > \sigma^2) \end{aligned}$$

Since $y_i \sim \mathcal{CN}(0, s_0\lambda_i + \sigma^2)$ and for large CNR (i.e., $s_0\lambda \gg \sigma^2$), $P_0 \approx 1$. Therefore as $s \to 0$, $g_i(s) < 0$ with probability close to one. Consequently, $f_i(s)$ is negative for small s but approaches

zero from the positive when $s \to \infty$. From equation (6.11), it is easy to see that there exists at least one solution for $f(s) = 0$.

In fact, a close inspection of (6.11) reveals that if the CNR is large for those dominant components (large λ_i's), existence of solutions is guaranteed with probability close to 1. To see this, notice that those terms in (6.11) with large λ_i dominate when $s \to 0$.

As to the uniqueness, while analytic proof has not been obtained, it is found through thorough numerical simulation that for large CNR, a unique solution is always determined. Fig. 6.1 is a typical example for $f(s)$ as a function of s.

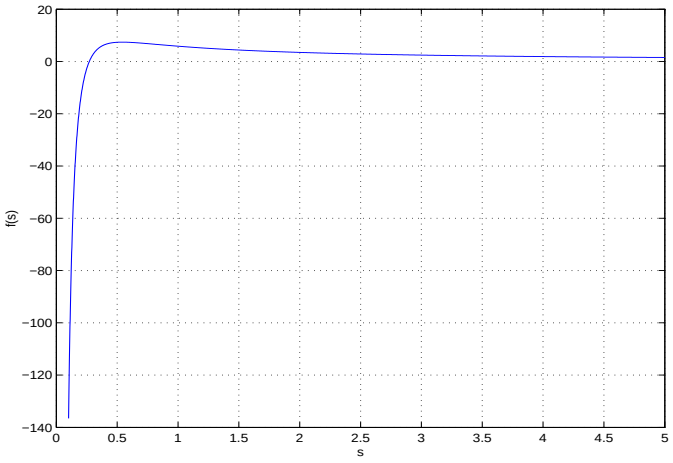


Figure 6.1: f(s) as a function of s.

6.4 Performance evaluation

6.4.1 Performance comparison with NMF

The ML estimates developed in the previous sections can be used to construct a GLRT for the detection problem specified in (6.1):

$$\begin{aligned} \Gamma_2 &= \frac{\max_{a,s,\sigma^2} f(\mathbf{x}|a,s,\sigma^2;\mathbf{H}_1)}{\max_{s,\sigma^2} f(\mathbf{x}|s,\sigma^2;\mathbf{H}_0)} \\ &= \frac{\max_{a,s,\sigma^2} \frac{1}{\|\mathbf{M}\|} \exp\left(-(\mathbf{x}-a\mathbf{v})^H \mathbf{M}^{-1}(\mathbf{x}-a\mathbf{v})\right)}{\max_{s,\sigma^2} \frac{1}{\|\mathbf{M}\|} \exp\left(-\mathbf{x}^H\mathbf{M}^{-1}\mathbf{x}\right)} \end{aligned} \tag{6.12}$$

In this section, we use numerical examples to compare the proposed GLRT with the NMF developed in [47, 48], which itself is a GLRT assuming clutter-only disturbance. In the first example, we use two channels, four pulses, hence $N = 8$, and the average $CNR = 40dB$. The output signal to interference and noise power ratio (SINR), defined as

$$SINR = 10\log_{10}|a|^2\mathbf{v}^H\mathbf{M}^{-1}\mathbf{v},$$

is fixed at 6dB. The clutter assumes a K distribution with a shape parameter $\alpha = 0.1$. The clutter ridge lies along the diagonal in the normalized Doppler-spatial frequency domain. The target signal is located at 0° azimuth and 0.15 normalized Doppler frequency in the spatial-temporal (Doppler) domain and the clutter has one lag temporal correlation $u_t = 0.999$ which helps determine the covariance matrix structure.

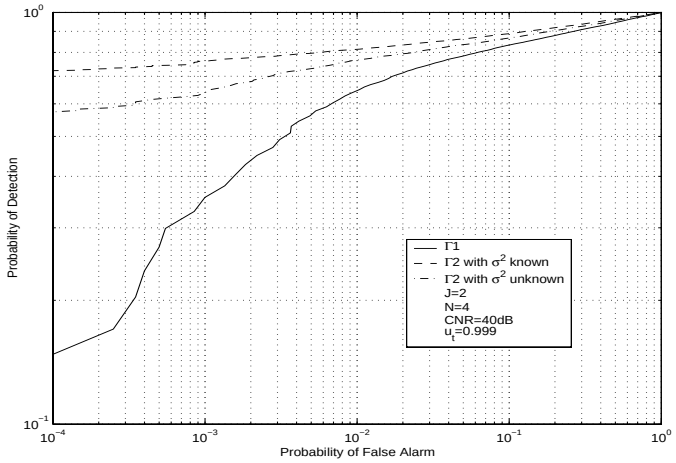


Figure 6.2: Performance comparison between the two GLRT (Γ_1 and Γ_2) in the presence of K distributed clutter and additive white Gaussian noise.

Fig. 6.2 gives the receiver operating characteristics (ROC) curves of the two statistics, namely the NMF and the proposed GLRT statistic. For the cases of both known σ^2 and unknown σ^2, the proposed GLRT of (6.12) outperforms the NMF of (6.2) by a significant margin. In the second example (shown in Fig. 6.3), we use a two channel thirty-two pulse example (hence $N = 64$) which is otherwise identical to the previous case. The same conclusion holds for this higher dimensioned case. The only difference is the curves for GLRT in the cases of known σ^2 and unknown σ^2 are almost identical. This is because of the improved estimation performance for σ^2 (and hence s as the nonlinear equations are coupled) in the higher dimension case due to the increased data size for each test cell.

Fig. 6.4 gives the probability of detection against the SINR for a fixed false alarm at 10^{-3} with $N = 8$ and σ^2 unknown. It can be easily seen that the proposed GLRT outperforms NMF, especially in the low SINR region.

6.4.2 Discussion of the CFAR property

In the absence of white Gaussian noise, the NMF of (6.2) has the desired CFAR property, i.e., the false alarm rate is independent of the clutter power term s. In the context of K distributed clutter, the CFAR with respect to power variation implies that it is CFAR with respect to the shape parameter.

In Fig. 6.5, the probability of false alarm as a function of the shape parameter is obtained via simulation for a threshold chosen such that the nominal false alarm rate is 10^{-3} in the clutter-only

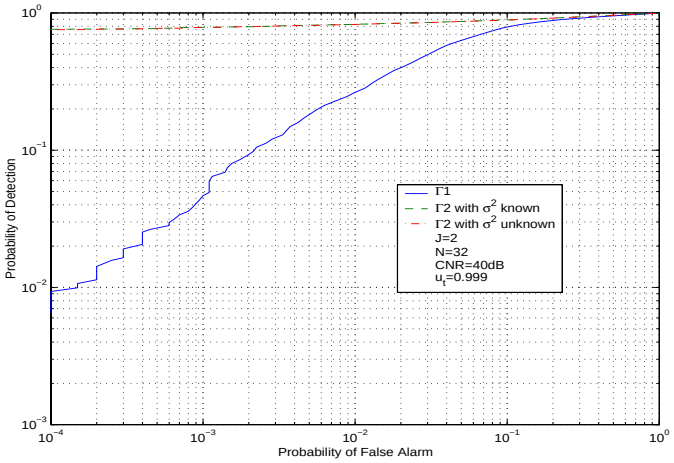


Figure 6.3: Same as in Fig. 6.2 except that $N = 64$ instead of 8.

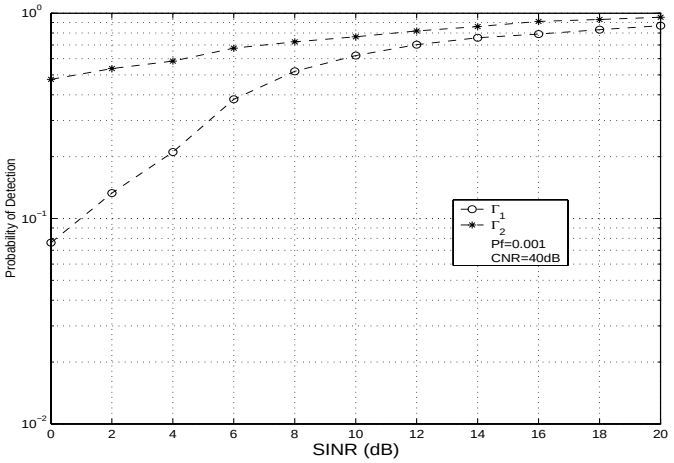


Figure 6.4: Probability of Detection as a function of SINR. $N = 8$ and σ^2 is unknown for Γ_2.

case. The average CNR is again fixed at $40dB$. Clearly, the false alarm rate changes significantly as a function of the shape parameter in the presence of additive noise, indicating the loss of CFAR for the NMF.

In Fig. 6.6, using the same setting as in the first example, the probability of false alarm of the proposed GLRT is given for a fixed threshold for the known σ^2 case. Notice that if the clutter texture term s is known perfectly, then the problem specified in (6.1) is a simple Gaussian noise problem with known covariance matrix $\mathbf{M}$ and the detection statistic in (6.12) reduces to the matched filter for Gaussian disturbance. Hence it is clearly CFAR with respect to s. The fact that we have to estimate s changes the CFAR property as shown in Fig. 6.6, most noticeably in the region with very small shape parameter. In this particular example, we notice that the proposed GLRT is still CFAR with respect to the shape parameter when it is greater than 0.1. The reason can be explained as follows. The ML estimate of s is likely to be very accurate for large CNR. At very

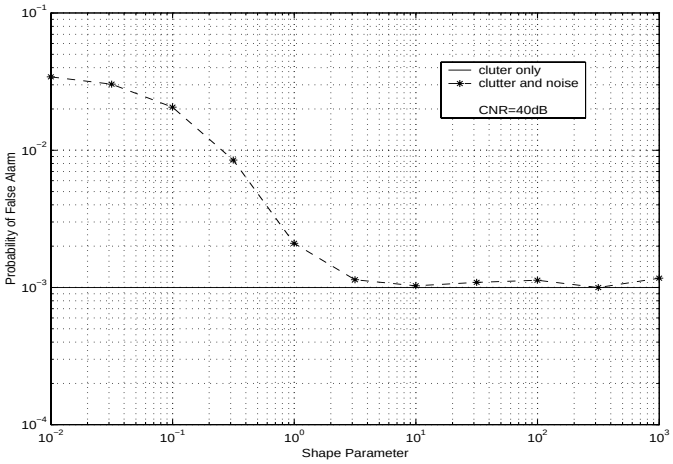


Figure 6.5: Flase alarm rate of the NMF statistic (Γ_1) as a function of the shape parameter of the K clutter. The nominal $P_f = 0.001$.

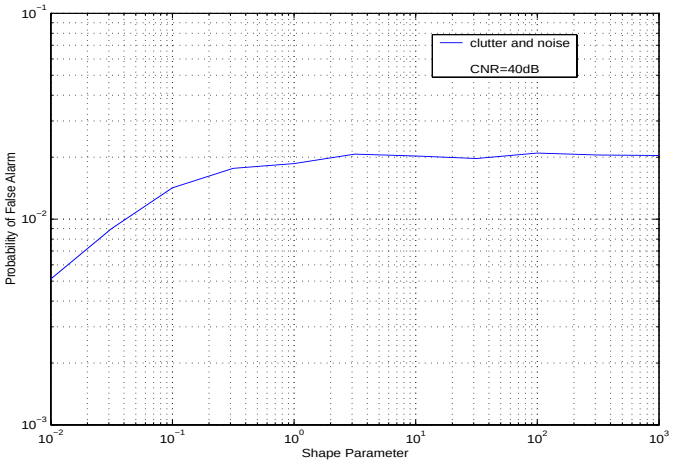


Figure 6.6: Flase alarm rate of the new GLRT (Γ_2) as a function of the shape parameter of the K clutter. The vector dimension $N = 8$.

low shape parameter values, the variance of the clutter texture term s becomes large. Therefore, even if the average CNR is kept at $40dB$, the likelihood of having smaller CNR increases. This results in a larger error variance of the estimate for s which in turn affects the CFAR property of the proposed statistic.

This CFAR performance will improve as the dimension N increases. Illustrated in Fig. 6.7 is a case for $N = 64$ that shows a better CFAR property than that of $N = 8$. This is due to the improved estimation performance for s for large N, as mentioned before. Notice that for the proposed GLRT, there is no nominal false alarm rate for a fixed threshold as the evaluation of false alarm probability is generally intractable.

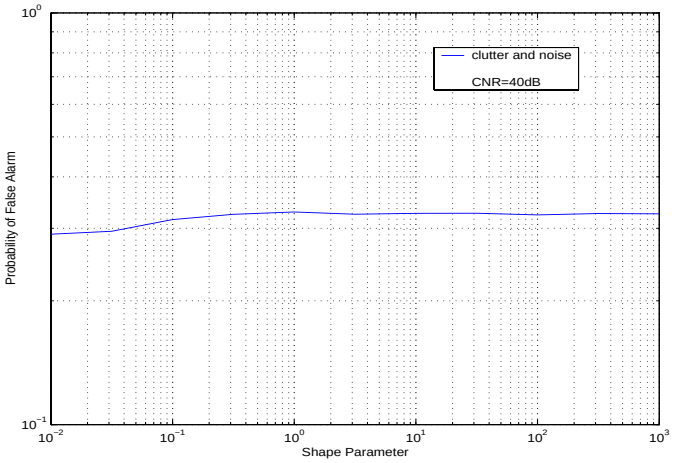


Figure 6.7: Same as in Fig 6.6 except that $N = 64$.

6.5 Conclusions

In this chapter, we consider the detection problem for the case of unknown, constant signal amplitude in the presence of non-Gaussian clutter plus additive white noise. The ML estimates of parameters associated with the detection problem, including both the clutter texture component and the target amplitude are derived. A simple bisection algorithm is devised for solving the ML equations. The developed estimates are then used to construct a GLRT test which can be shown to outperform the NMF developed for the clutter-only case, although at the expense of increased computational complexity. We also observe, through numerical examples, the NMF loses CFAR due to the presence of additive white noise while the proposed GLRT retains the CFAR property for a wide range of shape parameter values.

Chapter 7

Maximum Likelihood Estimation of Covariance Matrix for compound-Gaussian Processes

Compound-Gaussian processes have found important applications in modeling clutter returns for high-resolution radar. In this chapter we develop a maximum likelihood estimate for the covariance structure of a compound Gaussian process. The performance of the covariance matrix estimator is then evaluated in the context of adaptive radar detection. Through extensive numerical simulation and by using a popular CFAR detector for coherent pulse train detection in non-Gaussian clutter, we show that the proposed estimator provides better detection performance over existing covariance matrix estimators*.

7.1 Introduction

Experimental studies using high-resolution airborne radar clutter returns strongly indicate that the disturbance is no longer Gaussian. In particular, it was widely reported that the compound-Gaussian processes provide a better description to the statistical behavior of the power variation of clutter returns. In compound-Gaussian processes, the disturbance is modeled as a product of a real and non-negative scalar, s, and a correlated zero mean Gaussian vector, $\mathbf{c}$. Even though $\mathbf{c}$ is modeled as stationary complex Gaussian process, the variation of s from range cell to range cell renders the product process non-Gaussian. A particularly important compound-Gaussian process, the so-called spherically invariant random processes (SIRP) imposes a parametric model on s and therefore is amenable to analytical approaches [44–46]. Special cases of SIRP include the well studied K distributed clutter and Weibull clutter.

The observation of non-Gaussian disturbance has spurred great interest in space-time adaptive processing (STAP) for non-Gaussian clutter. A recently proposed statistic has been recognized as

*The materials presented in this chapter was never published in the open literature. Similar ML estimation procedure was independently derived by Conte, *et al*, and reported in [52].

a robust detection scheme for compound-Gaussian disturbance and exhibits a constant false alarm rate (CFAR) property with regard to the clutter power variation [47, 48]. Consider target detection under compound-Gaussian clutter, i.e., we want to distinguish the following two hypotheses given the test cell observation vector $\mathbf{x}$:

$$\begin{aligned} \mathbf{H}_0 \quad & \mathbf{x} = s\mathbf{c} \\ \mathbf{H}_1 \quad & \mathbf{x} = a\mathbf{v} + s\mathbf{c} \end{aligned} \tag{7.1}$$

where $\mathbf{c}$ is complex Gaussian with covariance matrix $\mathbf{\Sigma}$, a is the unknown signal amplitude, and $\mathbf{v}$ is the steering vector. The test statistic, derived as a generalized likelihood ratio test by treating s and a as unknown *constant* parameters [47], has the following form:

$$\Gamma = \frac{|\mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{v}|^2}{\left(\mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x}\right)\left(\mathbf{v}^H \mathbf{\Sigma}^{-1} \mathbf{v}\right)} \tag{7.2}$$

This proposed test statistic, herein termed as the normalized matched filter (NMF) due to the fact that it adds a normalizing constant to the well known matched filter detector for Gaussian disturbances, requires the knowledge of the covariance matrix structure of the compound-Gaussian process. Adaptive schemes that build on various covariance matrix estimators using secondary target-free data have since been proposed. While the sample covariance matrix (SCM) can be used as a heuristic estimate for the true covariance matrix, a more elegant estimator that tried to mitigate the effect due to the power variations was proposed [53–55]. Further, if the compound-Gaussian process is indeed an SIRP and assuming that the probability density function on the real and non-negative scalar component s is available, a maximum likelihood (ML) estimate of the covariance matrix can be obtained using the expectation maximization algorithm [56, 57].

In this chapter, we develop an ML estimator of the covariance matrix by assuming a general compound-Gaussian process. That is, we do not make any assumption about (or utilize) the statistics for the power term s. Instead, we treat the scalar variable s for the secondary data as an unknown constant conditioned on each realization or coherent processing interval (CPI) of the data. The developed algorithm therefore is not restricted to any particular SIRP processes. The performance of the estimator is evaluated using the detection performance by plugging the estimate in the NMF test statistic and compared with the SCM and the estimator proposed in [53, 54]. We show that it consistently provides the best detection performance among the three.

7.2 Maximum Likelihood Estimation of Covariance Matrix for Compound-Gaussian Processes

Assume that we have K secondary data, each of them is a $N \times 1$ vector with the k^{th} vector expressed as

$$\mathbf{z}_k = s_k \mathbf{c}_k \tag{7.3}$$

where $\mathbf{c}_k \sim CN(\mathbf{0}, \mathbf{\Sigma})$ and s_k's are some unknown real and non-negative constants. Our goal is to estimate $\mathbf{\Sigma}$ using the K training data $\mathbf{z}_k$'s. The log likelihood function can be written

straightforwardly as

$$\begin{aligned} L(\mathbf{\Sigma}, s_1, \cdots, s_K) &= \log f(\mathbf{z}_1, \cdots, \mathbf{z}_K | \mathbf{\Sigma}, s_1, \cdots, s_K) \\ &= -NK\log(\pi) - K\log|\mathbf{\Sigma}| - \sum_{k=1}^{K}\left[N\log(s_k) + \frac{1}{s_k}\mathbf{z}_k^H\mathbf{\Sigma}^{-1}\mathbf{z}_k\right] \end{aligned}$$

where $|\mathbf{A}|$ denotes the determinant of matrix $\mathbf{A}$ and $(\cdot)^H$ denotes the conjugate transpose of a matrix. We seek to maximize $L(\mathbf{\Sigma}, s_1, \cdots, s_K)$ with respect to $\mathbf{\Sigma}$ as well as s_k for $k = 1, \cdots, K$. Taking the derivative with respect to $\mathbf{\Sigma}$ and setting it equal to zero, we get

$$-K(\mathbf{\Sigma}^{-1})^T - \left[\mathbf{\Sigma}^{-1}\left(\sum_{k=1}^{K}\frac{1}{s_k}\mathbf{z}_k\mathbf{z}_k^H\right)\mathbf{\Sigma}^{-1}\right]^T = 0 \tag{7.4}$$

where $(\cdot)^T$ denotes the transpose of a matrix. In obtaining the above results, we have used the following facts

$$\begin{aligned} \frac{\partial \log|\mathbf{\Sigma}|}{\partial \mathbf{\Sigma}} &= \left(\mathbf{\Sigma}^{-1}\right)^T \\ \frac{\partial \mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z}}{\partial \mathbf{\Sigma}} &= \left(\mathbf{\Sigma}^{-1}\mathbf{z}\mathbf{z}^H\mathbf{\Sigma}^{-1}\right)^T \end{aligned}$$

These can be derived staightforwardly from the definition of derivatives with respect to matrices [58, 59]. From (7.4) it follows that:

$$\mathbf{\Sigma} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{s_k}\mathbf{z}_k\mathbf{z}_k^H \tag{7.5}$$

Taking the derivative of $L(\mathbf{\Sigma}, s_1, \cdots, s_K)$ with respect to s_k and set it equal to zero, we get

$$s_k = \frac{1}{N}\mathbf{z}_k^H\mathbf{\Sigma}^{-1}\mathbf{z}_k$$

Plug it back to (7.5), we have

$$\mathbf{\Sigma} = \frac{N}{K}\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\mathbf{\Sigma}^{-1}\mathbf{z}_k} \tag{7.6}$$

The solution to (7.6) yields the ML estimate of $\mathbf{\Sigma}$. Notice the subtle difference between the ML estimate and the estimate proposed in [53, 54] which has the form

$$\hat{\mathbf{\Sigma}} = \frac{N}{K}\sum_{k=1}^{K}\frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\mathbf{z}_k} \tag{7.7}$$

where the normalizing term is a simple inner product of the corresponding secondary data vector, as opposed to the quadratic term in the ML estimate in (7.6). We call the estimator in (7.7) herein as normalized sample covariance matrix (N-SCM) [55].

Notice that finding the ML estimate of $\mathbf{\Sigma}$ from (7.6) requires an iterative algorithm which is more complicated than either of SCM and N-SCM. However, the iteration is fairly straightforward to implement — it is in the same form as (7.6) by updating the current estimate of $\mathbf{\Sigma}$ from the previous one. Further, we found through numerical simulation that the algorithm usually converges after only a few (less than 8) iterations. To further expedite the convergence, initialization using SCM or N-SCM can be employed.

7.3 Performance Comparison

In this section, we compare the performance of the three covariance matrix estimators, namely the SCM, N-SCM, and ML estimators, in the context of the multichannel radar detection problem. We assume that the disturbance follows a K distribution that has been widely used to model clutter backscatter from airborne radars. The K distribution, a special case of SIRP, can be represented as a product of the square root of a Gamma distributed scalar random variable with a Gaussian vector. Using the model in (7.3) and assuming s_k^2 follows a Gamma distribution, the resulting $\mathbf{z}_k$ will have an amplitude that follows the K distribution with the probability density function

$$f(u) = \frac{b^{\alpha+1}u^{\alpha}}{2^{\alpha-1}\Gamma(\alpha)}K_{\alpha-1}(bu)$$

where $\Gamma(\cdot)$ is the Euler Gamma function and $K_a(\cdot)$ is the modified Bessel function of the second kind of order α. The parameter α is the shape parameter for the Gamma distributed clutter power term s^2. For K clutter, this shape parameter, controls the deviation from Gaussian disturbance. For example, as α approaches infinity, K distribution approaches the Gaussian distribution.

In the simulation, we choose $N = 16$ (pulse-channel product) with sample support $K = 50$. We evaluate the detection performance for various shape parameters using the three covariance matrix estimators, namely the ML estimator, SCM, and N-SCM. The output signal to noise ratio (SNR) is defined as [50]:

$$SNR = 10\log_{10}|a|^2\mathbf{v}^H\boldsymbol{\Sigma}^{-1}\mathbf{v}$$

Fig. 7.1 gives the probability of detection as a function of SNR for shape parameter $\alpha = 0.2$. In the figure we also plot the performance of the NMF test of (7.2) using the true covariance matrix which provides an upper bound on all three adaptive forms of the NMF. While numerically we can verify that ML covariance matrix is better that the other two covariance estimators, the advantage over the N-SCM is negligible in this case. In Fig. 7.2 we plot the probability of detection as a function of SNR for shape parameter $\alpha = 4$. The performance advantage of the ML estimator can be noticed. To further illustrate this, in Fig. 7.3, we fix the SNR at 10dB and plot the probability of detection against the shape parameter for the three covariance matrix estimators. Clearly from the figure, while the performance difference between ML and N-SCM at small shape parameter is negligible, as α gets larger, the performance advantage becomes noticeable. Therefore, the ML estimate provides some robustness on the detection performance with regard to the variation of the shape parameter.

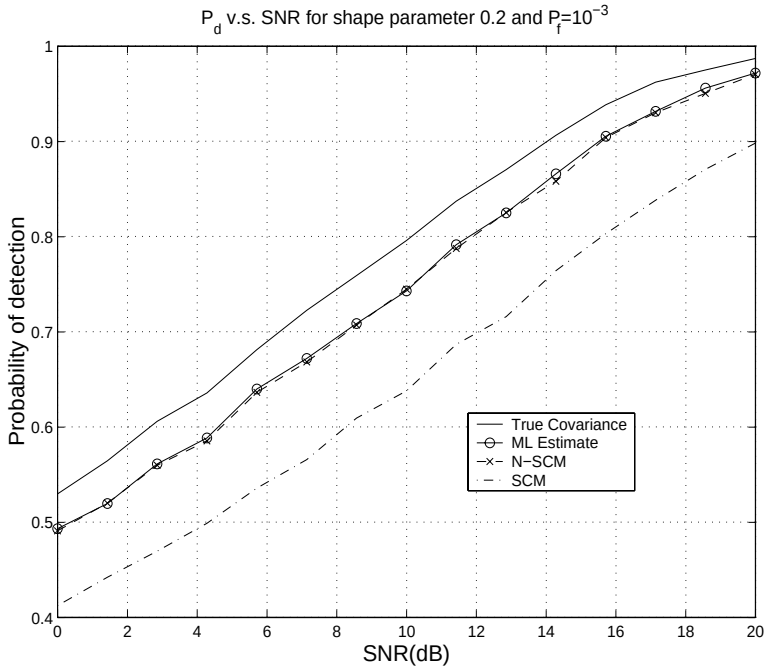


Figure 7.1: Probability of detection using the NMF as a function of SNR for different covariance matrix estimators. The shape parameter for the K clutter is 0.2 and the false alarm rate is 10^{-3}.

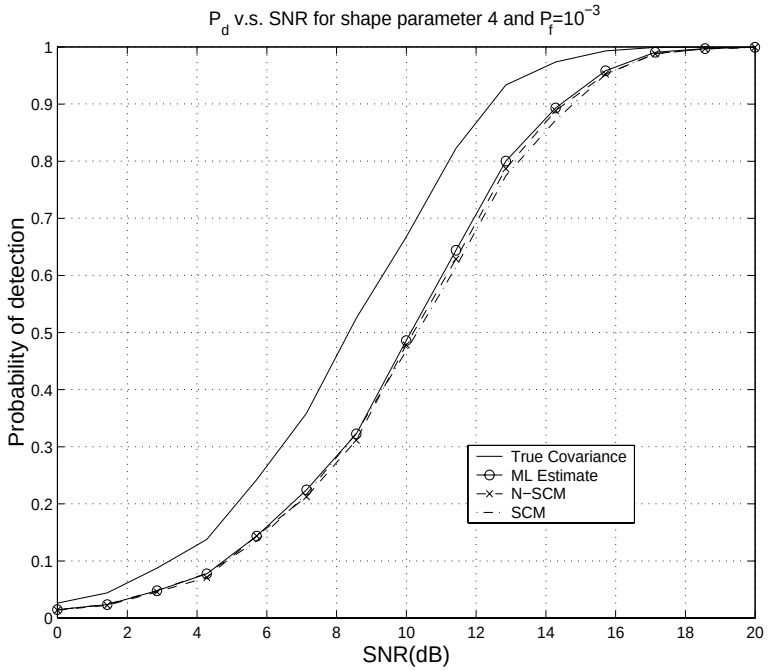


Figure 7.2: Probability of detection using the NMF as a function of SNR for different covariance matrix estimators. The shape parameter for the K clutter is 4 and the false alarm rate is 10^{-3}.

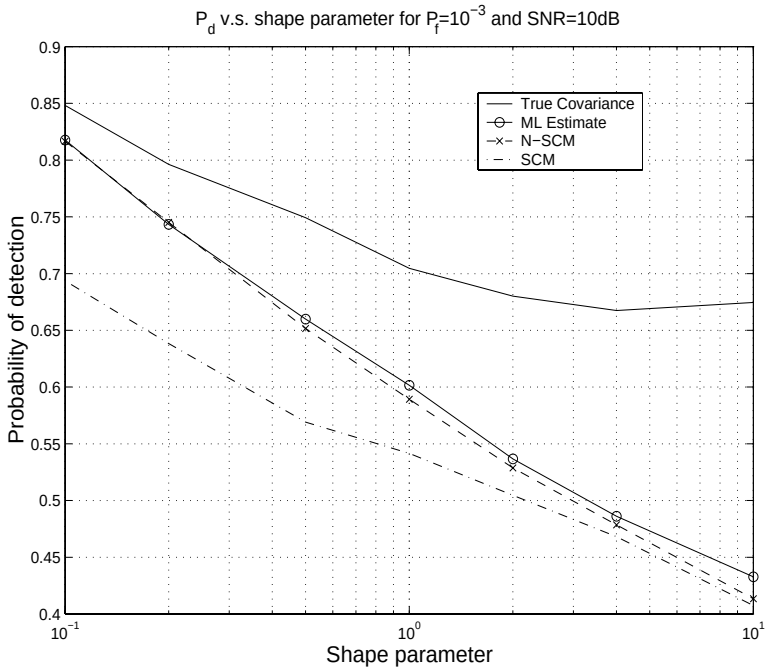


Figure 7.3: Probability of detection using the NMF as a function of shape parameter for different covariance matrix estimators. The SNR is 10dB and the false alarm rate is 10^{-3}.

Appendix A

List of Publications

Journal Publications

1. B. Liu, B. Chen, and J.H. Michels, "A GLRT for multichannel radar detection in the presence of both SIRP clutter and additive white Gaussian noise,", *submitted to the IEEE Trans. Signal Processing*, October, 2003, under revision.

2. B. Chen, P.K. Varshney, and J.H. Michels, "Adaptive CFAR Detection For Clutter-edge Heterogeneity Using Bayesian Inference," *IEEE Trans. Aerospace and Elec. Sys.*, vol. 39, no. 4, Oct. 2003.

3. T. Kasetkasem and P.K. Varshney, "Statistical characterization of clutter scenes based on a Markov random field model," *IEEE Trans. Aerospace and Electronic Systems*, vol. 39, pp. 1035 -1050, Jul. 2003.

4. B. Chen and P.K. Varshney, "A Bayesian sampling approach to decision fusion," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1809-1818, August 2002.

Refereed Conference Proceedings

1. T. Kasetkasem and P.K. Varshney, "Clutter patch identification based on Markov random field models" Proc. 2002 IEEE Radar Conference, Long Beach, CA, April 2002, pp. 464 -470.

2. B. Liu, B. Chen, and J.H. Michels, "A GLRT for radar detection in the presence of non-Gaussian clutter and additive white Gaussian noise," Proc. 2002 IEEE Sensor Array and Multichannel Signal Processing Workshop, Rosslyn, VA, Aug 2002, pp. 87-91.

3. B. Chen, P.K. Varshney, and J.H. Michels, "Adaptive CFAR detection via hierarchical model based parameter estimation," *Proceedings of the 35th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2001, pp. 1396–1400.

4. B. Chen, "Distribute detection using a Bayesian sampling approach," Proc. of 4th International Conference on Information Fusion, Montreal, Canada, August 2001.

Bibliography

[1] B.V. Dasarathy, *Decision Fusion*, IEEE Computer Society Press, Los Alamitos, CA, 1994.

[2] P.K. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, 1997.

[3] Edt. P.K. Varshney, "Special issue on data fusion," *Proceedings of the IEEE.*, vol. 85, no. 1, Jan. 1997.

[4] P.K. Willett, M. Alford, and V. Vannicola, "The case for like-sensor predetection fusion," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 30, pp. 986–1000, Oct. 1994.

[5] Z. Chair and P.K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerospace Elect. Sys.*, vol. 22, pp. 98–101, Jan. 1986.

[6] E. Drakopoulos and C.C. Lee, "Optimum multisensor fusion of correlated local decisions," *IEEE Trans. on Aerospace Elect. Syst.*, vol. 27, no. 4, pp. 593–605, July 1991.

[7] M. Kam, W. Chang, and Q. Zhu, "Hardware complexity of binary distributed detection systems with isolatd local Bayesian detectors," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 565–571, May/June 1991.

[8] M. Kam, Q. Zhu, and W.S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Trans. Aerospace Elect. Syst.*, vol. 28, pp. 916–920, July 1992.

[9] P.K. Willett, P.F. Swaszek, and R.S. Blum, "The good, bad, and ugly: distributed detection of a known signal in dependent Gaussian noise," *IEEE Trans. Signal Processing*, vol. 48, pp. 3266–3279, Dec. 2000.

[10] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, Wiley, New York, 1993.

[11] R.E. Kass and L. Wasserman, "The selection of prior distributions by formal rules," *Journal of the American Statistical Association*, vol. 91, pp. 1343–1370, Sep 1996.

[12] H. Jeffreys, *Theory of Probability*, Oxford University Press, London, UK, 3rd edition, 1961.

[13] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, Chapman and Hall Ltd, New York, 1995.

[14] A.E. Gelfand and A.F.M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, pp. 398–409, 1990.

[15] S. Ross, *Stochastic Processes*, Wiley, New York, 2nd edition, 1996.

[16] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

[17] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products*, Academic Press, San Diego, CA, 5th edition, 1994.

[18] Q. Yan and R. Blum, "Distributed signal detection under the Neyman-Pearson criterion," *IEEE Trans. Information Theory*, vol. 47, pp. 1368–1377, May 2001.

[19] L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, NY, 1986.

[20] B.D. Ripley, *Stochastic Simulation*, Wiley, New York, NY, 1987.

[21] G. Minkler, *CFAR: The Principles of Automatic Radar Detection in Clutter*, Magellan Book Co., 1990.

[22] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Trans. Aerospace and Electronic Systems*, vol. 19, pp. 608–621, July 1983.

[23] I. Ozgunes, P.P. Ghandi, and S.A. Kassam, "A variably trimmed mean CFAR radar detector," *IEEE Trans. Aeros. Elec. Systems*, vol. 28, pp. 1002–1014, 1992.

[24] R. Viswanathan and A. Eftekhari, "A selection and estimation test for multiple target detection," *IEEE Trans. Aeros. Elec. Systems*, vol. 28, pp. 509–519, 1992.

[25] M. Lops and P.K. Willett, "Ll-CFAR: A flexible robust alternative," *IEEE Trans. Aeros. Elec. Systems*, vol. 30, pp. 41–54, 1994.

[26] P.P. Gandhi and S.A. Kassam, "Analysis of CFAR processors in nonhomogeous background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 427–445, 1988.

[27] M. Smith and P. Varshney, "Intelligent CFAR processor based on data variability," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 36, pp. 837–847, July 2000.

[28] B. Chen, P.K. Varshney, and J.H. Michels, "Adaptive CFAR detection via hierarchical model based parameter estimation," in *Proc. of the 35th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2001, pp. 1396–1400.

[29] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Royal Statistical Society*, vol. 39, no. Ser. B, 1977.

[30] R. Streit and P. Willett, "Detection of random transient signals via hyperparameter estimation," *IEEE Trans. Signal Processing*, vol. 47, pp. 1823–1834, July 1999.

[31] M. L. Skolnik, *Introduction to Radar Systems*, McGraw Hill, Boston, MA, 3rd edition, 2001.

[32] J.A. Ritcey and J.L. Hines, "Performance of max-mean level detector with and without censoring," *IEEE Trans. Aerospace and Electronic Systems*, vol. 25, pp. 213–223, 1989.

[33] E. K. Al-Hussaini, "Performance of the greatest-of and censored greater-of detectors in multiple target environments," *IEE Proceedings, Part F: Communications, Radar and Signal Processing*, vol. 152, no. 3, pp. 193–198, June 1988.

[34] S. L. Wilson, "Two CFAR algorithms for interfering targets and nonhomogeneous clutter," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 29, pp. 57–72, Jan. 1993.

[35] M.A. Slamani, *A New Approach to Radar Detection Based on the Partitioning and Statistical Characterization of the Surveillance Volume*, Ph.D. thesis, Syracuse University, Syracuse, NY, May 1994.

[36] A. Ozturk and E.J. Dudewicz, "A New Statistical Goodness-of-Fit Test Based on Graphical Representation," *Biom. J.*, vol. 34, no. 4, pp. 403–427, 1992.

[37] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesion Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–41, 1984.

[38] P. Bremaud, *Markov Chains, Gibbs Field, Monte Carlo Simulation and Queues*, Springer-Verlag, New York, 1999.

[39] J. Marroquin, S.Mittter, and T. Poggio, "Probabilistic Solution of Ill-Posed Problem in Computer Vision," *Journal of the American Statistician Association*, vol. 82, no. 397, pp. 76–89, Mar. 1987.

[40] G. Winkler, *Image Analysis, Random Field and Dynamic Monte Carlo Methods*, Springer-Verlag, New York, 1995.

[41] J. Begas, "On the Statistical Analysis of Dirty Pictures," *J.R. Statist. Soc.*, vol. B 49, no. 3, pp. 259–302, 1986.

[42] H.L. Van Trees, *Detection, Estimation and Modulation Theory*, vol. 1, Wiley, New York, 1968.

[43] P.J. Green, "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[44] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant random process," *IEE Proc., part F: Radar, Sonar, and Navigation*, vol. 134, pp. 191–197, April 1987.

[45] M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Trans. Aerospace and Electronic Systems*, vol. 29, pp. 111–123, Jan 1983.

[46] K. Yao, "A representation theorem and its applications to spherically invariant random processes," *IEEE Trans. Information Theory*, vol. 19, pp. 600–608, 1973.

[47] F. Gini, "Sub-optimum coherent radar detection in a mixture of K-distributed and Gaussian clutter," *IEE Proc. F, Radar, Sonar and Navigation*, pp. 39–48, Feb. 1997.

[48] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-Gaussian clutter," *IEEE Trans. Aerospace Electron. Systems*, vol. 80, pp. 615–625, April 1995.

[49] L.L. Scharf and T.L. McWhorter, "Adaptive matched subspace detector and adaptive coherence estimators," in *Proc. 30th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1996.

[50] J.H. Michels, B. Himed, and M. Rangaswamy, "Performance of STAP tests in Gaussian and compound-Gaussian clutter," *Digital Signal Processing*, vol. 10, pp. 309–324, Oct. 2000.

[51] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[52] E. Conte, A. De Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Processing*, vol. 51, pp. 1908–1915, Aug. 2002.

[53] F. Gini, M. Greco, and L. Verrazzani, "Detection problem in mixed clutter environment as a Gaussian problem by adaptive preprocessing," *IEE Electronics Letters*, pp. 1189–1190, 1995.

[54] E. Conte, M. Lops, and G. Ricci, "Adaptive matched filter detection in spherically invariant noise," *IEEE Signal Processing Letters*, vol. 3, pp. 248–250, Aug 1996.

[55] F. Gini and J.H. Michels, "Performance analysis of two covariance matrix estimators in compound-Gaussian clutter," *IEE Proc., Radar, Sonar and Navigation*, pp. 133–140, June 1999.

[56] R.J. Little and D.B. Rubin, *Statistical analysis with missing data*, Wiley, New York, 1987.

[57] M. Rangaswamy and J.H. Michels, "Statistical analysis of the nonhomogeneity detector for non-Gaussian interference backgrounds," in *Proc. 2001 Defence Applications of Signal Processing (DASP) Workshop*, Barossa Valley Resort, South Australia, Sept 2001.

[58] L.L. Scharf, *Statistical Signal Processing — Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Boston, MA, 1990.

[59] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ., 1996.